

111

МИНОБРНАУКИ РОССИИ  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ  
Н.Г. ЧЕРНЫШЕВСКОГО»**

**Механико-математический факультет**

**СОГЛАСОВАНО**

заведующий кафедрой ДУиМЭ  
д.ф.-м.н., профессор

 С.И.Дудов  
"27" 09 2024 г.

**УТВЕРЖДАЮ**

председатель НМК механико-  
математического факультета,  
к.ф.-м.н., доцент

 Тышкевич С.В.  
"17" 09 2024 г.

**Фонд оценочных средств**  
текущего контроля и промежуточной аттестации по дисциплине

**Интеллектуальный анализ данных**

Направление подготовки магистратуры  
**09.04.03 - ПРИКЛАДНАЯ ИНФОРМАТИКА**

Профиль подготовки магистратуры  
**Прикладная информатика в экономике**

Квалификация (степень) выпускника  
**Магистр**

Форма обучения  
очная

Саратов, 2024

## 1. Результаты обучения по дисциплине «Интеллектуальный анализ данных»

Код и наименование компетенции	Код и наименование индикатора (индикаторов) достижения компетенции	Результаты обучения	Виды заданий и оценочных средств
<b>ОПК-3.</b> Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями.	1.1_М.ОПК-3. Обладает знаниями принципов, методов и средств анализа и структурирования профессиональной информации.	<b>Знать:</b> принципы, методы и средства анализа и структурирования профессиональной информации. <b>Уметь:</b> анализировать профессиональную информацию, выделять в ней главное.	<b>Собеседование</b>
	2.1_М.ОПК-3. Анализирует профессиональную информацию, выделяет в ней главное.	<b>Владеть:</b> навыками структурирования, оформления и представления профессиональной информации в виде аналитических обзоров с обоснованными выводами и рекомендациями.	<b>Контрольная работа</b>
	3.1_М.ОПК-3. Структурирует, оформляет и представляет профессиональную информацию в виде аналитических обзоров с обоснованными выводами и рекомендациями.		
<b>ПК-8.</b> Способность обосновывать подходы используемые в бизнес-анализе и руководстве бизнес-анализом	1.1_М.ПК-8. Грамотно определяет подходы к проведению бизнес-анализа, к работе с информацией бизнес-анализа, к работе с заинтересованными сторонами, к разработке различных типов требований, к работе с изменениями различных типов требований, к оценке эффективности работы по бизнес-анализу.	<b>Знать:</b> основные показатели эффективности бизнеса, которые позволяют определить уровень эффективности и финансовое состояние дел предприятия, а также, методы, которыми производится их анализ и расчет. <b>Уметь:</b> осуществлять анализ финансовой, организационной, технологической деятельности, с выводами о динамике роста,	<b>Собеседование</b>  <b>Контрольные задания</b>

		<p>перспективности и конкурентоспособности.</p> <p><b>Владеть:</b> навыками анализа предприятия или компании, сбора и коммуникации требований заинтересованных сторон, оценки и проверки правильности принятых решений.</p>	
	2.1_М.ПК-8. Планирует, организовывает и проводит встречи и обсуждения с заинтересованными сторонами.	<p><b>Знать:</b> основы теории межличностной и групповой коммуникации в деловом взаимодействии.</p> <p><b>Уметь:</b> планировать, организовывать и проводить встречи и обсуждения с заинтересованными сторонами.</p> <p><b>Владеть:</b> навыками разработки планов взаимодействия с заинтересованными сторонами, взаимодействия с заинтересованными сторонами и мониторинга заинтересованных сторон.</p>	<p><b>Собеседование</b></p> <p><b>не предусмотрено</b></p>
	3.1_М.ПК-8. Выявляет, регистрирует, анализирует и классифицирует риски и разрабатывать комплекс мероприятий по их минимизации. Оформляет результаты бизнес-анализа в соответствии с выбранными подходами.	<p><b>Знать:</b> основы теории конфликтов, теории управления рисками.</p> <p><b>Уметь:</b> представлять информацию бизнес-анализа различными способами и в различных форматах для обсуждения с заинтересованными сторонами.</p> <p><b>Владеть:</b> навыками управления рисками, обусловленными взаимодействием с заинтересованными</p>	<p><b>Собеседование</b></p> <p><b>Контрольные задания</b></p>

	Применяет информационные технологии в объеме, необходимом для целей бизнес-анализа.	сторонами.	
<b>ПК-13.</b> Способность использовать и развивать методы научных исследований и инструментария в области проектирования и управления информационными системами в прикладных областях	1.1_М.ПК-13. Грамотно использует информацию о: источниках информации, необходимой для профессиональной деятельности, современном отечественном и зарубежном опыте в профессиональной деятельности, стандартах и методиках оценки качества, стандартах и методиках оценки качества ресурсов ИТ, управлении активами ИТ и конфигурациями ИТ, способы определения потребностей в уровне качества ресурсов ИТ, основах управления качеством, инструментах и методах проведения, управлении качеством в проектах, методах и приемах формализации задач, методах и приемах алгоритмизации поставленных задач, программных продуктах для графического отображения алгоритмов стандартных алгоритмах и области их применения, языках формализации функциональных спецификаций,	<b>Знать:</b> стандарты оценки качества информационных систем, характеристики, описывающие качество программного обеспечения. <b>Уметь:</b> грамотно использовать информацию относительно комплекса условий, сопутствующих разработке и функционированию информационных систем, оценивать качество информационных систем. <b>Владеть:</b> методами и моделями оценки и измерения эффективности автоматизированных информационных систем.	<b>Собеседование</b>  <b>Контрольные задания</b>

	<p>методологиях разработки программного обеспечения, нотациях и программных продуктах для графического отображения алгоритмов, нормативных документах, определяющих требования к оформлению программного кода, возможностях типовой ИС, предметной области автоматизации, инструментах и методах выявления требований, технологии межличностной и групповой коммуникации в деловом, взаимодействии, основы конфликтологии архитектуре, устройстве и функционирование вычислительных систем, коммуникационном оборудовании, сетевых протоколах, основах современных операционных систем, основах современных систем управления базами данных, устройстве и функционирование современных ИС, современных стандартах информационного взаимодействия систем, программных</p>		
--	--	--	--

	средствах и платформах инфраструктуры информационных технологий организаций, системах классификации и кодирования информации, в том числе присвоение кодов документам и элементам справочников, отраслевой нормативная техническая документация.		
	2.1_М.ПК-13. Осуществляет моделирование бизнес-процессов в типовой ИС, разработку прототипа ИС на базе типовой ИС в соответствии с требованиями, разработку код ИС и баз данных ИС, тестирование разрабатываемого модуля ИС, интеграционное тестирование ИС на основе тест-планов..	<p><b>Знать:</b> основы теории бизнес-процессов в информационных системах, программирование и управление базами данных.</p> <p><b>Уметь:</b> описывать, разрабатывать и оптимизировать бизнес-процессы в информационных системах, разрабатывать и тестировать информационные системы.</p> <p><b>Владеть:</b> методами программирования, управления базами данных, интеграционного тестирования для проверки связи между компонентами, а также взаимодействия с различными частями системы.</p>	<p><b>Собеседование</b></p> <p><b>не предусмотрено</b></p>

## 2.Показатели оценивания планируемых результатов обучения

Семестр	Шкала оценивания			
	2	3	4	5
P				

3 семестр	<b>Не знает</b> способы сбора и обработки данных; методы анализа данных, необходимых для решения задач распознавания, классификации, кластеризации, прогнозирования.	<b>Слабо знает</b> способы сбора и обработки данных; методы анализа данных, необходимых для решения задач распознавания, классификации, кластеризации, прогнозирования.	<b>В целом знает</b> принципы и методы исследований; способы сбора и обработки данных; методы анализа данных, необходимых для решения задач распознавания, классификации, кластеризации, прогнозирования.	<b>Знает</b> научные принципы и методы исследований; способы сбора и обработки данных; методы анализа данных, необходимых для решения задач распознавания, классификации, кластеризации, прогнозирования.
	<b>Не умеет</b> пользоваться методами интеллектуального анализа данных для анализа и решения прикладных задач распознавания, классификации, кластеризации, прогнозирования.	<b>Практически не умеет</b> пользоваться методами интеллектуального анализа данных для анализа и решения прикладных задач распознавания, классификации, кластеризации, прогнозирования.	<b>Умеет</b> использовать методы визуализации экспериментальных данных; выбирать инструментальные средства для обработки данных в соответствии с поставленной задачей; собирать и обрабатывать данные с помощью различных статистических методов; анализировать и интерпретировать статистические данные о социально-экономических явлениях и процессах;	<b>Умеет</b> использовать методы визуализации экспериментальных данных; выбирать инструментальные средства для обработки данных в соответствии с поставленной задачей; собирать и обрабатывать данные с помощью различных статистических методов; анализировать и интерпретировать статистические данные о социально-экономических явлениях и процессах; на практике применять новые научные принципы и методы исследований;
	<b>Не владеет</b> практическими навыками по использованию инструментальных средств анализа данных различной природы	<b>Слабо владеет</b> практическими навыками по использованию инструментальных средств анализа данных различной природы	<b>Владеет</b> практическими навыками по использованию инструментальных средств анализа данных различной природы.	<b>Уверенно владеет</b> практическими навыками по использованию инструментальных средств анализа данных различной природы Демонстрирует навыки

				самостоятельного решения поставленной задачи по поиску и обработке данных, построению модели и анализу и интерпретации результатов. Грамотно обосновывает свои решения и анализирует результаты расчетов.
--	--	--	--	---

## *Оценочные средства*

### 1.1 Задания для текущего контроля

#### **1) Задания для оценки «ОПК-3»**

##### *Собеседование*

##### *Вопросы для собеседования*

- Задачи анализа данных. Интеллектуальный анализ данных. Отличия от других видов анализа данных.
- Постановка и порядок решения задачи интеллектуального анализа данных на примере информационной модели торговой компании.
- Закономерности в данных: неочевидные, объективные, полезные.
- Области применения интеллектуального анализа данных.
- Подготовка данных для анализа. Обработка пропущенных и недостающих данных. Анализ исключений.
- Использование баз данных в интеллектуальном анализе данных.
- Постановка и порядок решения задачи интеллектуального анализа
- Особенности базы данных для оперативной аналитической обработки данных.
- Виды и особенности шкал измерений данных.
- Постановка и порядок решения задачи интеллектуального анализа
- данных на примере документной базы данных.
- Особенности использования данных при интеллектуальном анализе данных. Данные, информация и знания.
- Основные задачи интеллектуального анализа данных.
- Стадии интеллектуального анализа данных.

##### **Контрольная работа**

**Методические рекомендации.** Учебным планом по направлению подготовки 09.04.03 - "Прикладная Информатика" предусмотрена две контрольных работы. Каждая контрольная работа предполагает проведение компьютерных расчетов по заданной теме с использованием пакета R или электронных таблиц в зависимости от условия задания. Студенту требуется продемонстрировать умение работать с исходными данными; обосновывать выбор методов обработки данных; проводить вариантные расчеты и давать содержательную интерпретацию результатов. Решение контрольного задания

оформляется в форме отчета в письменном виде. Контрольная работа оценивается на основании теоретической аккуратности, логичности и аргументированности процесса анализа данных, полноты и проработанности исследования, качества интерпретации результатов и сделанных выводов.

Подготовка к контрольной работе осуществляется в период лекционных и практических занятий, а также во внеаудиторные часы в рамках самостоятельной работы. Во время самостоятельной подготовки студент пользуется конспектами аудиторных занятий, основной и дополнительной литературой по дисциплине (см. перечень литературы в рабочей программе дисциплины).

**Критерии оценивания.** Количество баллов, выставляемых за выполнение заданий, зависит от полноты решения и правильности ответа. Общие требования к выполнению заданий: решение должно быть грамотным, полным. За решение, в котором обосновано получен правильный ответ, выставляется максимальное количество баллов. Правильный ответ при отсутствии текста решения оценивается в 0 баллов.

Имеется верное обоснование утверждения и обоснованно получен верный ответ - 10 баллов.

Допущена единичная ошибка, возможно, приведшая к неверному ответу, но при этом имеется верная последовательность всех шагов решения - 5 баллов.

Решение не соответствует ни одному из критериев, перечисленных выше - 0 баллов.

### **Примеры вариантов для контрольной работы**

#### **Вариант 1**

По данным ExZ1.xls провести иерархическую классификацию объектов

1. Постройте диаграмму рассеивания. Предложите свой вариант классификации.
2. Пусть объекты 1-50 – относятся к первому классу, 51-100 – ко второму. Найдите расстояние между кластерами по принципу:
  - а) ближнего соседа
  - б) дальнего соседа
  - в) центров тяжести групп
  - г) средней связи
3. Постройте дерево классификации, измеряя расстояние между кластерами по принципу «ближнего соседа». В качестве расстояния примите обычное евклидово расстояние.
4. Постройте дерево классификации, измеряя расстояние между кластерами по принципу «средней связи». В качестве расстояния примите обычное евклидово расстояние.
5. Проведите классификацию объектов по методу k-средних
6. Сравните различные разбиения по критерию суммы внутриклассовых дисперсий.

#### **Вариант 2**

Проанализируйте данные о видовом составе 72 лесных делянок

*Данные: wood.xls*

1. Сравните результаты иерархического кластерного анализа использовании различных агломерационных алгоритмов. Какой метод кажется вам предпочтительным? Какое число кластеров вы выделили? Сколько объектов в каждом классе?
2. Проведите кластерный анализ методом K-средних. Сравните результаты классификации для разного числа классов.
3. Проверьте насколько хорошо разделяются классы с помощью дискриминантного анализа
4. Визуализируйте результаты классификации на координатной плоскости X-Y.

## ***Задания для оценки «ПК-8»***

### ***Собеседование*** ***Вопросы для собеседования***

Вычисление и анализ показателей описательной статистики.

Построение и использование моделей для интеллектуального анализа данных.

Решение задач кластеризация в ходе интеллектуального анализа данных: цель и основные методы.

Основные алгоритмы иерархической кластеризации.

Кластеризация методом k-средних и его обобщения, нечетких k-средних, k-медиоидов.

Алгоритмы кластеризации основанные на модели конечной смеси распределений. EM-алгоритм. MCLUST.

Спектральные методы кластеризации объектов.

Методы кластеризации основанные на анализе плотности. Алгоритмы DBSCAN, GDBSCAN, HDBSCAN, Optics и др.

Решение задач классификации в ходе интеллектуального анализа данных: цель и основные методы.

Метод k ближайших соседей.

Линейный дискриминантный анализ, квадратичный дискриминантный анализ.

Деревья решений.

Ансамбли моделей. Бэггинг. Бустинг.

Альтернативные методы построения ансамблей моделей.

### **Примеры вариантов для контрольной работы**

**Методические рекомендации.** Учебным планом по направлению подготовки 09.04.03 - "Прикладная Информатика" предусмотрена две контрольных работы. Каждая контрольная работа предполагает проведение компьютерных расчетов по заданной теме с использованием пакета R или электронных таблиц в зависимости от условия задания. Студенту требуется продемонстрировать умение работать с исходными данными; обосновывать выбор методов обработки данных; проводить вариантовые расчеты и давать содержательную интерпретацию результатов. Решение контрольного задания оформляется в форме отчета в письменном виде. Контрольная работа оценивается на основании теоретической аккуратности, логичности и аргументированности процесса анализа данных, полноты и проработанности исследования, качества интерпретации результатов и сделанных выводов.

Подготовка к контрольной работе осуществляется в период лекционных и практических занятий, а также во внеаудиторные часы в рамках самостоятельной работы. Во время самостоятельной подготовки студент пользуется конспектами аудиторных занятий, основной и дополнительной литературой по дисциплине (см. перечень литературы в рабочей программе дисциплины).

**Критерии оценивания.** Количество баллов, выставляемых за выполнение заданий, зависит от полноты решения и правильности ответа. Общие требования к выполнению заданий: решение должно быть грамотным, полным. За решение, в котором обоснованно получен правильный ответ, выставляется максимальное количество баллов. Правильный ответ при отсутствии текста решения оценивается в 0 баллов.

Имеется верное обоснование утверждения и обоснованно получен верный ответ - 10 баллов.

Допущена единичная ошибка, возможно, приведшая к неверному ответу, но при этом имеется верная последовательность всех шагов решения - 5 баллов.

**Решение не соответствует ни одному из критериев, перечисленных выше - 0 баллов.**

### **Вариант 1**

Файл *Customer.xls* содержит данные о стоимости первого купленного автомобиля (зависимая переменная), возрасте, поле, уровне дохода, образовании и семейном положении (независимые переменные) клиентов крупной сети магазинов по продаже автомобилей.

Используя алгоритм CRT построения дерева решений:

- построить модель, которая предсказывала бы стоимость приобретаемого автомобиля, если известны значения независимых переменных;
- Оценить качество модели;
- Применить построенную модель для предсказания стоимости покупаемого автомобиля клиентами магазина.
- 

### **Вариант 2**

В файле *CreditScore.xls* содержатся данные о кредитных рейтингах клиентов банка (результативная переменная), пользовавшихся банковскими услугами в прошлом, а также такие данные о клиентах, как возраст; уровень дохода; число используемых кредитных карт; уровень образования и число кредитов на покупку автомашины (нет или 1, 2 и более).

- Сформируйте обучающую выборку, отобрав случайным образом 70 % клиентов банка и 30% -ю тестовую выборку.
- Поведите дискриминантный анализ для объектов, включенных в обучающую выборку. Найдите значения коэффициентов дискриминантной функции, проверьте их статистическую значимость
- Проведите дискриминантный анализ для случая равных априорных вероятностей и априорных вероятностей, пропорциональных объему групп. Проверьте также, выполняется ли предположение о многомерной нормальности данных в группах.

### **Вариант 3**

В файле *NBAplayer.xls* приведены данные о заработной палате игроков NBA, физических характеристиках, амплуа, игровом времени и показателях результативности

- Используя метод К-средних для кластеризации по физическим характеристикам. Позволяют ли результаты кластеризации предсказывать амплуа игроков?
- Сравните результаты кластеризации по физическим характеристикам, полученных различными методами иерархической классификации;
- Постройте дискриминантные функции разделяющие игроков по амплуа;
- Используйте алгоритм случайного леса для исследования зависимости заработной платы баскетболиста от амплуа и результативности

## **Задания для оценки «ПК-13»**

### **Собеседование**

### **Вопросы для собеседования**

Процесс анализа данных в пакете R.

Импорт данных в среду R.

Библиотеки на языке R, предназначенные для решения задач кластерного анализа.

Библиотеки пакета R, предназначенные для решения задач классификации.

Графические возможности пакета R.

Среда разработки программного обеспечения RSTUDIO.

Возможности языка разметки Rmarkdown.

Краткий синтаксис Markdown и расширенный Markdown

Выделение блока кода

Ссылка на изображение и загрузка изображения

Математическая формула LaTex

Диаграммы последовательности и блок-схемы UML  
Ведение блога оффлайн  
Импорт и экспорт файлов Markdown  
Использование интерфейса Kaggle для отладки кода и подготовки отчетных документов.

## **1.2) Промежуточная аттестация**

*Методические указания.* Промежуточная аттестация по дисциплине проводится в виде устного экзамена. Учебным планом по направлению подготовки 09.04.03 - "Прикладная Информатика" предусмотрена одна промежуточная аттестация. Подготовка студента к прохождению промежуточной аттестации осуществляется в период аудиторных занятий, а также во внеаудиторные часы в рамках самостоятельной работы. Во время самостоятельной подготовки студент пользуется конспектами аудиторных занятий, основной и дополнительной литературой по дисциплине (см. перечень литературы).

*Критерии оценивания.* Во время экзамена студент должен дать развернутый ответ на вопросы, изложенные в билете и выполнить практическое задание. Преподаватель вправе задавать дополнительные вопросы по всему изучаемому курсу.

Во время ответа студент должен продемонстрировать знания современных методов и алгоритмов анализа данных, в том числе, методов визуализации экспериментальных данных; байесовской классификации, кластерного анализа, поиска ассоциативных правил и построения деревьев решений, выбор системы информативных признаков и оценивания качества алгоритмов, обосновывать выбора инструментальных средств для обработки данных в соответствии с поставленной задачей.

Полнота ответа определяется показателями оценивания планируемых результатов обучения.

### **Примерные вопросы к зачету**

Модели и их свойства. Аналитический и информационный походы к моделированию.	ОПК-3
Формы представления, типы и виды анализируемых данных.	ОПК-3
Обучение моделей «с учителем» и «без учителя». Обучающее и тестовое множество. Ошибки обучения. Эффект переобучения.	ОПК-3
Общая схема анализа данных. Требования к алгоритмам анализа данных.	ОПК-3
Data Mining. Характеристика классов задач, решаемых методами Data Mining.	ОПК-3
Программный инструментарий для выполнения анализа данных.	ОПК-3
Модели Data Mining. Сфера применения технологии Data Mining.	ОПК-3
Методы Data Mining: классификация. Основные алгоритмы классификационного анализа.	ОПК-3
Методы Data Mining: поиск ассоциативных правил. Основные алгоритмы ассоциативного анализа.	ОПК-3
Методы Data Mining: кластеризация. Основные алгоритмы кластерного анализа.	ОПК-3
Визуальный анализ данных (Visual Mining): этапы, преимущества и недостатки.	ОПК-3
Цели, задачи и основное содержание кластерного анализа. Классификация методов кластеризации.	ОПК-3
Способы определения меры расстояния между кластерами.	ОПК-3
Характеристика методов связи для процедуры кластеризации (одиночная, полная, средняя).	ОПК-3
Алгоритм кластеризации k-means.	ОПК-3

Алгоритмы кластеризации основанные на модели конечной смеси распределений. ЕМ-алгоритм. MCLUST.	ОПК-3
Спектральные методы кластеризации объектов.	ОПК-3
Методы кластеризации основанные на анализе плотности. Алгоритмы DBSCAN, GDBSCAN, HDBSCAN, Optics и др.	ОПК-3
Валидация и сравнение результатов работы различных алгоритмов кластеризации, выбор оптимального числа и состава кластеров	ОПК-3
Решение задач классификации в ходе интеллектуального анализа данных: цель и основные методы.	ПК-8
Метод k ближайших соседей.	ПК-8
Наивный байесовский классификатор	ПК-8
Логистическая регрессия	ПК-8
Линейный дискриминантный анализ, квадратичный дискриминантный анализ.	ПК-8
Деревья решений.	ПК-8
Цели, задачи и принципы построения деревьев решений. Общая характеристика алгоритмов построения деревьев решений.	ПК-8
Сфера применения деревьев решений.	ПК-8
Упрощение деревьев решений.	ПК-8
Алгоритм случайного леса.	ПК-8
Ансамбли моделей. Бэггинг. Бустинг.	ПК-8
Альтернативные методы построения ансамблей моделей.	ПК-8
Библиотеки на языке R, предназначенные для решения задач кластерного анализа.	ПК-13
Библиотеки пакета R, предназначенные для решения задач классификации.	ПК-13
Среда разработки программного обеспечения RSTUDIO.	ПК-13
Возможности языка разметки Rmarkdown.	ПК-13

### Контрольные задания

Подготовьте письменный отчет о проведении анализа предложенного набора данных Рекомендуемые для самостоятельного анализа наборы данных приведены в таблицах. Примечание. Если Вы хотите выбрать для самостоятельной работы другой набор данных, то это возможно, но необходимо заранее согласовать с преподавателем.

N	Источник данных	Набор данных
1.	<a href="https://sci2s.ugr.es/keel/category.php?cat=clas">https://sci2s.ugr.es/keel/category.php?cat=clas</a>	<a href="#">glass</a>
2.	<a href="https://sci2s.ugr.es/keel/category.php?cat=clas">https://sci2s.ugr.es/keel/category.php?cat=clas</a>	<a href="#">saheart</a>
3.	<a href="https://sci2s.ugr.es/keel/category.php?cat=clas">https://sci2s.ugr.es/keel/category.php?cat=clas</a>	<a href="#">vowel</a>
4.	<a href="https://sci2s.ugr.es/keel/category.php?cat=clas">https://sci2s.ugr.es/keel/category.php?cat=clas</a>	<a href="#">winequality-red</a>
5.	<a href="https://sci2s.ugr.es/keel/category.php?cat=clas">https://sci2s.ugr.es/keel/category.php?cat=clas</a>	<a href="#">winequality-white</a>
6.	<a href="https://sci2s.ugr.es/keel/category.php?cat=clas">https://sci2s.ugr.es/keel/category.php?cat=clas</a>	<a href="#">australian</a>
7.	<a href="https://sci2s.ugr.es/keel/category.php?cat=clas">https://sci2s.ugr.es/keel/category.php?cat=clas</a>	<a href="#">crx</a>
8.	<a href="https://sci2s.ugr.es/keel/category.php?cat=clas">https://sci2s.ugr.es/keel/category.php?cat=clas</a>	<a href="#">penbased</a>
9.	<a href="https://sci2s.ugr.es/keel/category.php?cat=clas">https://sci2s.ugr.es/keel/category.php?cat=clas</a>	<a href="#">automobile</a>
10.	<a href="https://sci2s.ugr.es/keel/category.php?cat=clas">https://sci2s.ugr.es/keel/category.php?cat=clas</a>	<a href="#">spambase</a>
11.	<a href="https://sci2s.ugr.es/keel/category.php?cat=clas">https://sci2s.ugr.es/keel/category.php?cat=clas</a>	<a href="#">german</a>
12.	<a href="https://sci2s.ugr.es/keel/category.php?cat=clas">https://sci2s.ugr.es/keel/category.php?cat=clas</a>	<a href="#">wine</a>
13.	<a href="https://sci2s.ugr.es/keel/category.php?cat=clas">https://sci2s.ugr.es/keel/category.php?cat=clas</a>	<a href="#">coil2000</a>
14.	<a href="https://sci2s.ugr.es/keel/category.php?cat=clas">https://sci2s.ugr.es/keel/category.php?cat=clas</a>	<a href="#">marketing</a>

15.	<a href="https://sci2s.ugr.es/keel/category.php?cat=clas">https://sci2s.ugr.es/keel/category.php?cat=clas</a>	<a href="#">housevotes</a>
16.	<a href="https://sci2s.ugr.es/keel/category.php?cat=clas">https://sci2s.ugr.es/keel/category.php?cat=clas</a>	<a href="#">thyroid</a>
17.	<a href="https://sci2s.ugr.es/keel/category.php?cat=clas">https://sci2s.ugr.es/keel/category.php?cat=clas</a>	<a href="#">spambase</a>
18.	<a href="https://sci2s.ugr.es/keel/category.php?cat=clas">https://sci2s.ugr.es/keel/category.php?cat=clas</a>	<a href="#">Census</a>
19.	<a href="https://sci2s.ugr.es/keel/category.php?cat=clas">https://sci2s.ugr.es/keel/category.php?cat=clas</a>	<a href="#">penbased</a>
20.	<a href="https://sci2s.ugr.es/keel/category.php?cat=clas">https://sci2s.ugr.es/keel/category.php?cat=clas</a>	<a href="#">automobile</a>
21.	<a href="https://sci2s.ugr.es/keel/category.php?cat=clas">https://sci2s.ugr.es/keel/category.php?cat=clas</a>	<a href="#">spambase</a>
22.	<a href="https://sci2s.ugr.es/keel/category.php?cat=clas">https://sci2s.ugr.es/keel/category.php?cat=clas</a>	<a href="#">german</a>
23.	<a href="https://sci2s.ugr.es/keel/category.php?cat=clas">https://sci2s.ugr.es/keel/category.php?cat=clas</a>	<a href="#">wine</a>
24.	<a href="https://sci2s.ugr.es/keel/category.php?cat=clas">https://sci2s.ugr.es/keel/category.php?cat=clas</a>	<a href="#">coil2000</a>

Отчет должен содержать:

1. Описание набора данных. (**компетенция ОПК-3**)
  - 1.1. Постановку задачи. Поясните, какая переменная является целевой, смысл прочих переменных
  - 1.2. Описательные статистики переменных, графики распределений, диаграммы рассеивания с выделением классов и т.д.
2. Формулировка задачи кластеризации наблюдений (целевая переменная не используется) (**компетенция ПК-8**)
  - 2.1. Решение задач кластеризации методами:
    - 2.1.1. Иерархической кластеризации (методами ближнего соседа, дальнего соседа, средней связи, Уорда). Поясните какие признаки были выбраны, способ вычисления расстояния между объектами, сколько классов вы выделили, состав классов.
    - 2.1.2. Методом k-средних, нечетких k-средних (kmeans, kmedoids, cmeans)
    - 2.1.3. Кластеризация на основе модели смеси распределений (Mclust, pam)
    - 2.1.4. Спектральных методов (kernlab, Spectrum)
    - 2.1.5. Анализа плотностей (Dbscan)
    - 2.1.6. Другие алгоритмы и библиотеки, по вашему выбору
  - 2.2. Визуализацию кластерных решений
  - 2.3. Выполните валидацию и сравнение результатов работы различных алгоритмов, рекомендации по выбору алгоритма, оптимального числа и состава кластеров (clValid, clusterSim и т.п.).
3. Решение задач классификации (прогнозируемую переменную выберете в соответствии с п.1) (**компетенция ПК-13**)
  - 3.1. Разделите данные на обучающую и контрольную выборку или используйте наборы данных для кросс-валидации.
  - 3.2. Сравните результаты применения различных алгоритмов классификации:
    - 3.3. Линейный дискриминантный анализ ( lda )
    - 3.4. Квадратичный дискриминантный анализ ( qda )
    - 3.5. Логистическая регрессия ( mlogit, ologit )
    - 3.6. Метод k-ближайших соседей ( knn )
    - 3.7. Наивный байесовский классификатор ( Naïve bayes )
    - 3.8. Деревья решений ( rpart, party )
    - 3.9. Ансамбли моделей:
      - 3.9.1. Случайный лес ( randomforest )

- 3.9.2. Бэггинг (ipred::bagging и т.п.)
- 3.9.3. Бустинг (adabag, gbm, XGBOOST и т.п.)
- 3.10. Сравните точность полученных решений и сформулируйте выводы о выборе решающих правил

ФОС для проведения промежуточной аттестации одобрен на заседании кафедры дифференциальных уравнений и математической экономики, протокол № 4 от 27 сентября 2024 г.

Автор

Профессор, д.э.н., профессор Балаш В.А.Балаш