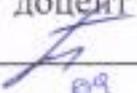


МИНОБРНАУКИ РОССИИ  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ  
Н.Г. ЧЕРНЫШЕВСКОГО»

**Механико-математический факультет**

СОГЛАСОВАНО  
заведующий кафедрой ДУиМЭ  
д.ф.-м.н., профессор  
  
С.И.Дудов  
" 28 " 08 2024 г.

УТВЕРЖДАЮ  
председатель НМК механико-  
математического факультета,  
к.ф.-м.н., доцент  
  
Тышкевич С.В.  
" 27 " 08 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  
**Анализ данных**

Квалификация (степень) выпускника  
**Магистр**

Форма обучения  
очная

Саратов, 2024

## 1. Результаты обучения по дисциплине «Спецкурс 11.1»

Код и наименование компетенции	Код и наименование индикатора (индикаторов) достижения компетенции	Результаты обучения	Виды заданий и оценочных средств
<p><b>ПК-5</b> Способен собирать, систематизировать, обрабатывать числовую и не числовую информацию. Подготавливать аналитические обзоры и рекомендации на основе различных математических расчетов.</p>	<p>1.1_М.ПК-5 Выбирает методы и математические модели решения для проведения различных математических (в том числе статистических и актуарных) расчетов.</p>	<p><b>Знать:</b> - математические методы и модели, применяемые для решения задач кластеризации и классификации объектов; <b>Уметь:</b> - применять методы решения задач кластеризации и классификации объектов; <b>Владеть:</b> - методами решения задач кластеризации и классификации объектов;</p>	<p>Собеседование</p> <p>Задания для практических и лабораторных занятий</p> <p>Контрольная работа</p>
	<p>2.1_М.ПК-5 Собирает информацию, анализирует достаточность и качество предоставленной информации для проведения расчетов.</p>	<p><b>Знать:</b> - методологию сбора, обработки и анализа данных <b>Уметь:</b> - выбирать средства и методы, достаточные для решения задач сбора, анализа и обработки данных - осуществлять проведение работ по сбору и обработке информации; - оценивать качество полученной информации; - анализировать достаточность полученной информации. <b>Владеть:</b> - навыками сбора и обработки информации, формирования базы данных для задач</p>	<p>Собеседование</p> <p>Задания для практических и лабораторных занятий</p>

		анализа временных рядов и прогнозирования	
	3.1_М.ПК-5 Производит различные отчеты с помощью современных информационных технологий и программирования. На основании которых готовит отчет и дает рекомендации по изменению стратегии управления организации.	<b>Знать:</b> - способы организации отчетов с помощью современных информационных технологий и программирования. <b>Уметь:</b> - составлять отчеты по результатам работы составленных алгоритмов. <b>Владеть:</b> - навыками публичного представления полученных результатов.	Собеседование  Задания для практических и лабораторных занятий
<b>ПК - 6</b> Способен анализировать требования к программному обеспечению, системные требования к базам данных, планировать управление требованиями, изменениями и аналитическими работами в ИТ проекте.	1.1_М.ПК-6 Анализирует возможности и оценивает трудоемкость реализации требований к программному обеспечению, базам данных и ИТ проектам.	<b>Знать:</b> - способы оценки трудоемкость реализации требований к программному обеспечению, интеллектуального анализа данных. <b>Уметь:</b> - планировать управление требованиями, изменениями и аналитическими работами в ИТ проекте <b>Владеть:</b> - навыками анализа требований к программному обеспечению, управления аналитическими работами в ИТ проекте	Собеседование  Задания для практических и лабораторных занятий
	2.1_М.ПК-6 Понимает принципы работы, технологии и возможности современных и перспективных средств разработки программного продукта, баз данных и ИТ проектов.		
	3.1_М.ПК-6 Вырабатывает варианты реализации всевозможных требований и изменений.		

<p><b>ПК – 7</b> Способен проектировать и разрабатывать инструменты и методы для реализации бизнес-процессов в ИТ проектах.</p>	<p>1.1_М.ПК-7 Строит математические модели бизнес-процессов предприятия с учетом проведенного анализа.</p>	<p><b>Знать:</b> - методологию моделирования бизнес-процессов, место подсистемы прогнозирования в ИТ проектах <b>Уметь:</b> - использовать современные системы моделирования бизнес-процессов <b>Владеть:</b> - навыками анализа бизнес-процессов предприятия</p>	<p>Собеседование</p> <p>Задания для практических и лабораторных занятий</p>
	<p>2.1_М.ПК-7 Создает инструменты и методы разработки ИТ проекта и адаптирует его к бизнес-процессам предприятия.</p>	<p><b>Знать:</b> - типовые инструменты и методы анализа и прогнозирования бизнес-процессов <b>Уметь:</b> - разрабатывать инструменты и методы анализа данных и прогнозирования и адаптировать их к бизнес-процессам предприятия <b>Владеть:</b> - навыками создания инструментов и методов анализа данных и прогнозирования при разработке ИТ проекта и их адаптации к бизнес-процессам предприятия</p>	<p>Собеседование</p> <p>Задания для практических и лабораторных занятий</p>
	<p>3.1_М.ПК-7 Реализует полученную модель бизнес-процесса с использованием ИТ.</p>	<p><b>Знать:</b> - информационные системы, библиотеки и модули, реализующие методы анализа и прогнозирования бизнес-процессов <b>Уметь:</b> - использовать стандартные методы, библиотеки и модули анализа временных</p>	<p>Собеседование</p> <p>Задания для практических</p>

		рядов и прогнозирования при реализации полученных моделей бизнес-процессов с использованием ИТ <b>Владеть:</b> - навыками реализации построенных моделей бизнес-процессов с использованием ИТ.	и лабораторных занятий
	4.1_М.ПК- 7 Анализирует необходимость модернизации программного обеспечения, баз данных и информационных систем и проводит ее в соответствии с полученными требованиями к изменениям.	<b>Знать:</b> - основные типы программного обеспечения и информационных систем, применяемых при решении задач и анализа временных рядов и прогнозирования бизнес-процессов <b>Уметь:</b> - анализировать необходимость модернизации программного обеспечения, баз данных и информационных систем, применяемых при решении задач анализа временных рядов и прогнозирования бизнес-процессов <b>Владеть:</b> - навыками модернизации программного обеспечения обработки, анализа данных и прогнозирования бизнес-процессов в соответствии с требованиями	Собеседование  Задания для практических и лабораторных занятий
<b>ПК – 9</b> Способен собирать, обрабатывать, систематизировать числовую и нечисловую информацию, проводить анализ и на его основе делать выводы о ситуации	1.1_М.ПК-9 Разрабатывает план проведения работ по анализу бизнес-процессов.	<b>Знать:</b> - основные этапы проведения работ по анализу бизнес-процессов. <b>Уметь:</b> - разрабатывать план проведения работ по	Собеседование

в организации.		<p>анализу бизнес-процессов.</p> <p><b>Владеть:</b></p> <p>- навыками разработки плана анализа бизнес-процессов.</p>	<p>Задания для практических и лабораторных занятий</p>
	<p>2.1_М.ПК-9</p> <p>Определяет и обосновывает комплекс мероприятий по оптимизации работы организации.</p>	<p><b>Знать:</b></p> <p>- принципы обоснования комплекса мероприятий по оптимизации работы организации.</p> <p><b>Уметь:</b></p> <p>- определять комплекс мероприятий по оптимизации работы организации, применяя методы теории случайных процессов и анализа временных рядов.</p> <p><b>Владеть:</b></p> <p>- методами теории случайных процессов и временных рядов и навыками применения этих методов для оптимизации работы организации.</p>	<p>Собеседование</p> <p>Задания для практических и лабораторных занятий</p>
	<p>3.1_М.ПК-9</p> <p>Принимает решение по определению целей и задач стратегических изменений в организации.</p>	<p><b>Знать:</b></p> <p>- основные этапы математического моделирования при принятии решений о стратегических изменениях в организации.</p> <p><b>Уметь:</b></p> <p>- принимать решение по определению целей и задач стратегических изменений в организации, применяя методы теории случайных процессов и временных рядов.</p> <p><b>Владеть:</b></p> <p>- навыками разработки стратегии достижения</p>	

		цели с использованием методов теории случайных процессов и временных рядов.	
<p><b>ПК – 13</b> Способен разрабатывать и совершенствовать различные методы анализа массовых количественных данных, разрабатывать систему математического обеспечения решения задач, возникающих в процессе проведения расчетов, представлять научные результаты на конференциях в виде докладов и публикаций.</p>	<p>1.1_М.ПК-13 Производит различные математические (в том числе статистические и актуарные) расчеты на основе соответствующих математических и технических средств, в том числе с помощью пакета прикладных программ.</p>	<p><b>Знать:</b> - математические методы, модели, инструментальные средства и пакеты прикладных программ анализа временных рядов и прогнозирования <b>Уметь:</b> - выбирать и использовать инструментальные средства и пакеты прикладных программ анализа временных рядов и прогнозирования <b>Владеть:</b> - навыками программирования и реализации методов анализа временных рядов и прогнозирования с помощью пакетов прикладных программ (R, Python).</p>	<p>Собеседование</p> <p>Задания для практических и лабораторных занятий</p>
	<p>2.1_М.ПК-13 Совершенствует математические методы анализа.</p>	<p><b>Знать:</b> - основные методы теории анализа временных рядов и прогнозирования. <b>Уметь:</b> - корректно применять и совершенствовать методы теории эконометрического моделирования и анализа временных рядов при решении задач прогнозирования. <b>Владеть:</b> - информацией о современном состоянии и развитии методов анализа временных рядов в мире.</p>	<p>Собеседование</p> <p>Задания для практических и лабораторных занятий</p>
	<p>3.1_М.ПК-13 Проводит</p>	<p><b>Знать:</b> - способы организации</p>	<p>Собеседование</p>

	<p>презентации полученных и новых результатов, обосновывает актуальность и эффективность работы новых методик.</p>	<p>отчетов с помощью современных информационных технологий и программирования.</p> <p><b>Уметь:</b></p> <ul style="list-style-type: none"> <li>- использовать современные информационные технологии и программирования для представления полученных и новых результатов;</li> <li>- обосновывать актуальность и эффективность работы новых методик.</li> </ul> <p><b>Владеть:</b></p> <ul style="list-style-type: none"> <li>- навыками публичного представления полученных результатов на конференциях в виде докладов и публикаций.</li> </ul>	<p>Задания для практических и лабораторных занятий</p>
--	--	--	--

## 2. Показатели оценивания планируемых результатов обучения

Семестр	Шкала оценивания			
	2	3	4	5
3 семестр	<p><b><u>Не знает</u></b> способы сбора и обработки данных; методы анализа данных, необходимых для решения задач распознавания, классификации, кластеризации, прогнозирования.</p> <p><b><u>Не умеет</u></b> пользоваться методами интеллектуального анализа данных для анализа и решения</p>	<p><b><u>Слабо знает</u></b> способы сбора и обработки данных; методы анализа данных, необходимых для решения задач распознавания, классификации, кластеризации, прогнозирования.</p> <p><b><u>Практически не умеет</u></b> пользоваться методами интеллектуального анализа данных для анализа и</p>	<p><b><u>В целом знает</u></b> принципы и методы исследований; способы сбора и обработки данных; методы анализа данных, необходимых для решения задач распознавания, классификации, кластеризации, прогнозирования.</p> <p><b><u>Умеет</u></b> использовать методы визуализации экспериментальных данных; выбирать инструментальные средства для обработки данных в</p>	<p><b><u>Знает</u></b> научные принципы и методы исследований; способы сбора и обработки данных; методы анализа данных, необходимых для решения задач распознавания, классификации, кластеризации, прогнозирования.</p> <p><b><u>Умеет</u></b> использовать методы визуализации экспериментальных данных; выбирать инструментальные средства для обработки данных в</p>

	<p>прикладных задач распознавания, классификации, кластеризации, прогнозирования.</p> <p><b><u>Не владеет</u></b> практическими навыками по использованию инструментальных средств анализа данных различной природы</p>	<p>решения прикладных задач распознавания, классификации, кластеризации, прогнозирования.</p> <p><b><u>Слабо владеет</u></b> практическими навыками по использованию инструментальных средств анализа данных различной природы</p>	<p>соответствии с поставленной задачей; собирать и обрабатывать данные с помощью различных статистических методов; анализировать и интерпретировать статистические данные социально-экономических явлениях и процессах;</p> <p><b><u>Владеет</u></b> практическими навыками по использованию инструментальных средств анализа данных различной природы.</p>	<p>соответствии с поставленной задачей; собирать и обрабатывать данные с помощью различных статистических методов; анализировать и интерпретировать статистические данные социально-экономических явлениях и процессах; на практике применять новые научные принципы и методы исследований;</p> <p><b><u>Уверенно владеет.</u></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.</p>
--	---	--	---	---

### **1.1 Задания для текущего контроля**

1) ПК-5

***Собеседование  
Вопросы для собеседования***

Задачи анализа данных. Интеллектуальный анализ данных. Отличия от других видов анализа данных.  
Закономерности в данных: неочевидные, объективные, полезные.  
Области применения интеллектуального анализа данных.  
Подготовка данных для анализа. Обработка пропущенных и недостающих данных.  
Анализ исключений.  
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.  
Основные задачи интеллектуального анализа данных.  
Стадии интеллектуального анализа данных.

### ***Задания для оценки «ПК-6»***

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

Использование баз данных в интеллектуальном анализе данных.  
Постановка и порядок решения задачи интеллектуального анализа  
Особенности базы данных для оперативной аналитической обработки данных.  
Концепции и средства управления данными в R.  
Возможности библиотеки data.table.  
Инструментарий пакета tidyvers. Формат организации данных tibble.  
Возможности библиотеки dplyr для организации процесса обработки данных. Конвейеры.

### ***Задания для оценки «ПК-7»***

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

Возможности анализа бизнес-процессов в пакете R  
Библиотека vupaR.  
Извлечение информации из журнала событий.  
Описательный анализ частоты событий. Построение диаграммы частот действий на основе событий.  
Построение и кастомизация карты процессов.  
Отображение показателей производительности на карте процессов  
Построение матрица прецедентов.  
Анализ ресурсов и ативностей.

### ***Задания для оценки «ПК-9»***

#### ***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-медоидов.

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.

Библиотеки пакета R, предназначенные для решения задач кластерного анализа.

Библиотеки пакета R, предназначенные для решения задач классификации.

#### 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 варианты расчетов 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

1. Используя метод K-средних для кластеризации по физическим характеристикам. Позволяют ли результаты кластеризации предсказывать амплуа игроков?
2. Сравните результаты кластеризации по физическим характеристикам, полученных различными методами иерархической классификации;
3. Постройте дискриминантные функции разделяющие игроков по амплуа;
4. 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.

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.

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, выбора 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 подходы к моделированию.	ПК-5
Формы представления, типы и виды анализируемых данных.	ПК-5
Обучение моделей «с учителем» и «без учителя». Обучающее и тестовое множество. Ошибки обучения. Эффект переобучения.	ПК-5
Общая схема анализа данных. Требования к алгоритмам анализа данных.	ПК-5
Data Mining. Характеристика классов задач, решаемых методами Data Mining.	ПК-5
Программный инструментарий для выполнения анализа данных.	ПК-5
Модели Data Mining. Сферы применения технологии Data Mining.	ПК-5
Использование баз данных в интеллектуальном анализе данных.	ПК-6
Возможности библиотеки data.table пакета R.	ПК-6
Инструментарий библиотеки tidyvers. Формат организации данных tibble.	ПК-6
Возможности библиотеки dplyr для организации процесса обработки данных.	ПК-6
Методы Data Mining: классификация. Основные алгоритмы классификационного анализа.	ПК-9
Методы Data Mining: поиск ассоциативных правил. Основные алгоритмы ассоциативного анализа.	ПК-9
Методы Data Mining: кластеризация. Основные алгоритмы кластерного анализа.	ПК-9
Визуальный анализ данных (Visual Mining): этапы, преимущества и недостатки.	ПК-9
Цели, задачи и основное содержание кластерного анализа. Классификация методов кластеризации.	ПК-9
Способы определения меры расстояния между кластерами.	ПК-9
Характеристика методов связи для процедуры кластеризации (одиночная,	ПК-9

полная, средняя).	
Алгоритм кластеризации k-means.	ПК-9
Алгоритмы кластеризации основанные на модели конечной смеси распределений. EM-алгоритм. MCLUST.	ПК-9
Спектральные методы кластеризации объектов.	ПК-9
Методы кластеризации основанные на анализе плотности. Алгоритмы DBSCAN, GDBSCAN, HDBSCAN, Optics и др.	ПК-9
Валидация и сравнение результатов работы различных алгоритмов кластеризации, выбор оптимального числа и состава кластеров	ПК-9
Решение задач классификации в ходе интеллектуального анализа данных: цель и основные методы.	ПК-9
Метод k ближайших соседей.	ПК-9
Наивный байесовский классификатор	ПК-9
Логистическая регрессия	ПК-9
Линейный дискриминантный анализ, квадратичный дискриминантный анализ.	ПК-9
Деревья решений.	ПК-9
Цели, задачи и принципы построения деревьев решений. Общая характеристика алгоритмов построения деревьев решений.	ПК-9
Сферы применения деревьев решений.	ПК-9
Упрощение деревьев решений.	ПК-9
Алгоритм случайного леса.	ПК-9
Ансамбли моделей. Бэггинг. Бустинг.	ПК-9
Альтернативные методы построения ансамблей моделей.	ПК-9
Возможности анализа бизнес-процессов в пакете R	ПК-7
Библиотека vupaR. Извлечение информации из журнала событий.	ПК-7
Библиотека vupaR. Построение и кастомизация карты процессов.	ПК-7
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
Среда разработки программного обеспечения RSTUDIO.	ПК-13
Возможности языка разметки Rmarkdown.	ПК-13
Использование интерфейса Kaggle для отладки кода и подготовки отчетных документов.	ПК-13

### Контрольные задания

Подготовьте письменный отчет о проведении анализа предложенного набора данных. Рекомендуются для самостоятельного анализа наборы данных приведены в таблицах. Для оформления отчета используйте язык разметки Rmarkdown или интерфейс Kaggle. 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. Постановку задачи. Поясните, какая переменная является целевой, смысл прочих переменных
  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**)
  1. Решение задач кластеризации методами:
    1. Иерархической кластеризации (методами ближнего соседа, дальнего соседа, средней связи, Уорда). Поясните какие признаки были выбраны, способ вычисления расстояния между объектами, сколько классов вы выделели, состав классов.
    2. Методом k-средних, нечетких k-средних (kmeans, kmedoids, smean)
    3. Кластеризация на основе модели смеси распределений (Mclust, pam)
    4. Спектральных методов (kernlab, Spectrum)
    5. Анализа плотностей (DbSCAN)
    6. Другие алгоритмы и библиотеки, по вашему выбору
  2. Визуализацию кластерных решений
  3. Выполните валидацию и сравнение результатов работы различных алгоритмов, рекомендации по выбору алгоритма, оптимального числа и состава кластеров (cValid, clusterSim и т.п.).
3. Решение задач классификации (прогнозируемую переменную выберете в соответствии с п.1) (**компетенция ПК-13**)
  1. Разделите данные на обучающую и контрольную выборку или используйте наборы данных для кросс-валидации.

2. Сравните результаты применения различных алгоритмов классификации:
3. Линейный дискриминантный анализ (lda)
4. Квадратичный дискриминантный анализ (qda)
5. Логистическая регрессия (mlogit, ologit)
6. Метод k-ближайших соседей (knn)
7. Наивный байесовский классификатор (Naïve bayes)
8. Деревья решений (rpart, party)
9. Ансамбли моделей:
  1. Случайный лес (randomforest)
  2. Бэггинг (ipred::bagging и т.п.)
  3. Бустинг (adabag, gbm, XGBOOST и т.п.)
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