

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное учреждение
высшего профессионального образования
"Казанский (Приволжский) федеральный университет"
Институт вычислительной математики и информационных технологий



УТВЕРЖДАЮ

Проректор
по образовательной деятельности КФУ
Проф. Минзарипов Р.Г.

_____ 20__ г.

Программа дисциплины

Научно-исследовательский семинар МЗ.Б.3

Направление подготовки: 010400.68 - Прикладная математика и информатика

Профиль подготовки: Математическое и информационное обеспечение экономической деятельности

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Автор(ы):

-

Рецензент(ы):

Миссаров М.Д.

СОГЛАСОВАНО:

Заведующий(ая) кафедрой:

Протокол заседания кафедры No ____ от " ____ " _____ 201__ г

Учебно-методическая комиссия Института вычислительной математики и информационных технологий:

Протокол заседания УМК No ____ от " ____ " _____ 201__ г

Регистрационный No

Казань
2014

Содержание

1. Цели освоения дисциплины
2. Место дисциплины в структуре основной образовательной программы
3. Компетенции обучающегося, формируемые в результате освоения дисциплины /модуля
4. Структура и содержание дисциплины/ модуля
5. Образовательные технологии, включая интерактивные формы обучения
6. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины и учебно-методическое обеспечение самостоятельной работы студентов
7. Литература
8. Интернет-ресурсы
9. Материально-техническое обеспечение дисциплины/модуля согласно утвержденному учебному плану

Программу дисциплины разработал(а)(и)

1. Цели освоения дисциплины

Научно-исследовательский семинар в 3 семестре посвящен интеллектуальному анализу данных.

Курс охватывает следующие разделы:

- Предварительная обработка данных;
- Задача классификации с обучением;
- Поиск ассоциативных правил;
- Кластерный анализ;
- Нечеткая логика и нечеткие множества.

2. Место дисциплины в структуре основной образовательной программы высшего профессионального образования

Данная учебная дисциплина включена в раздел " М3.Б.3 Научно-исследовательская работа магистра" основной образовательной программы 010400.68 Прикладная математика и информатика и относится к базовой (общепрофессиональной) части. Осваивается на 1, 2 курсах, 1, 2, 3 семестры.

Данный курс проводится на 4 курсе, в 7 семестре.

3. Компетенции обучающегося, формируемые в результате освоения дисциплины /модуля

В результате освоения дисциплины формируются следующие компетенции:

В результате освоения дисциплины студент:

1. должен знать:

о современных методах Data Mining.

2. должен уметь:

понимать основные проблемы, возникающие при анализе данных, и пути их решения.

3. должен владеть:

навыками анализа данных различной природы.

4. Структура и содержание дисциплины/ модуля

Общая трудоемкость дисциплины составляет зачетных(ые) единиц(ы) 216 часа(ов).

Форма промежуточного контроля дисциплины зачет в 1 семестре; зачет во 2 семестре; зачет в 3 семестре.

Суммарно по дисциплине можно получить 100 баллов, из них текущая работа оценивается в 50 баллов, итоговая форма контроля - в 50 баллов. Минимальное количество для допуска к зачету 28 баллов.

86 баллов и более - "отлично" (отл.);

71-85 баллов - "хорошо" (хор.);

55-70 баллов - "удовлетворительно" (удов.);

54 балла и менее - "неудовлетворительно" (неуд.).

4.1 Структура и содержание аудиторной работы по дисциплине/ модулю

Тематический план дисциплины/модуля

N	Раздел Дисциплины/ Модуля	Семестр	Неделя семестра	Виды и часы аудиторной работы, их трудоемкость (в часах)			Текущие формы контроля
				Лекции	Практические занятия	Лабораторные работы	
1.	Тема 1. Введение в интеллектуальный анализ данных. Области использования. Связь с другими дисциплинами. Компоненты систем Data Mining. Основные этапы анализа данных. Исторический обзор	3	1-3	0	0	4	
2.	Тема 2. Предварительная обработка данных. Очистка данных. Интеграция и преобразование данных. Сокращение данных.	3	4-6	0	0	6	
3.	Тема 3. Задача классификации с обучением. Классификация с использованием деревьев решений, нейронных сетей. "Наивная" байесовская классификация, байесовские сети. Метод k ближайших соседей.	3	7-9	0	0	6	
4.	Тема 4. Задача поиска ассоциативных правил на примере анализа рыночной корзины. Метод ?A priori? генерации ассоциативных правил.	3	10-12	0	0	6	
5.	Тема 5. Кластерный анализ. Типы данных в кластерном анализе. Методы k средних, k медоидов. Агломеративные и дивизимные методы иерархической кластеризации.	3	13-15	0	0	6	

N	Раздел Дисциплины/ Модуля	Семестр	Неделя семестра	Виды и часы аудиторной работы, их трудоемкость (в часах)			Текущие формы контроля
				Лекции	Практические занятия	Лабораторные работы	
6.	Тема 6. Введение в нечеткую логику. Определение нечетких множеств. Основные операции с нечеткими множествами. Алгоритм не-четкой кластеризации.	3	16-18	0	0	6	
	Тема . Итоговая форма контроля	1		0	0	0	зачет
	Тема . Итоговая форма контроля	2		0	0	0	зачет
	Тема . Итоговая форма контроля	3		0	0	0	зачет
	Итого			0	0	34	

4.2 Содержание дисциплины

Тема 1. Введение в интеллектуальный анализ данных. Области использования. Связь с другими дисциплинами. Компоненты систем Data Mining. Основные этапы анализа данных. Исторический обзор

лабораторная работа (4 часа(ов)):

Введение в интеллектуальный анализ данных. Области использования. Связь с другими дисциплинами. Компоненты систем Data Mining. Основные этапы анализа данных. Исторический обзор

Тема 2. Предварительная обработка данных. Очистка данных. Интеграция и преобразование данных. Сокращение данных.

лабораторная работа (6 часа(ов)):

Предварительная обработка данных. Очистка данных. Интеграция и преобразование данных. Сокращение данных.

Тема 3. Задача классификации с обучением. Классификация с использованием деревьев решений, нейронных сетей. "Наивная" байесовская классификация, байесовские сети. Метод k ближайших соседей.

лабораторная работа (6 часа(ов)):

Задача классификации с обучением. Классификация с использованием деревьев решений, нейронных сетей. "Наивная" байесовская классификация, байесовские сети. Метод k ближайших соседей.

Тема 4. Задача поиска ассоциативных правил на примере анализа рыночной корзины. Метод ?A priori? генерации ассоциативных правил.

лабораторная работа (6 часа(ов)):

Задача поиска ассоциативных правил на примере анализа рыночной корзины. Метод ?A priori? генерации ассоциативных правил.

Тема 5. Кластерный анализ. Типы данных в кластерном анализе. Методы k средних, k медоидов. Агломеративные и дивизимные методы иерархической кластеризации.

лабораторная работа (6 часа(ов)):

5 Кластерный анализ. Типы данных в кластерном анализе. Методы k средних, k медоидов. Агломеративные и дивизимные методы иерархической кластеризации.

Тема 6. Введение в нечеткую логику. Определение нечетких множеств. Основные операции с нечеткими множествами. Алгоритм не-четкой кластеризации.

лабораторная работа (6 часа(ов)):

Введение в нечеткую логику. Определение нечетких множеств. Основные операции с нечеткими множествами. Алгоритм не-четкой кластеризации.

4.3 Структура и содержание самостоятельной работы дисциплины (модуля)

N	Раздел Дисциплины	Семестр	Неделя семестра	Виды самостоятельной работы студентов	Трудоемкость (в часах)	Формы контроля самостоятельной работы
1.	Тема 1. Введение в интеллектуальный анализ данных. Области использования. Связь с другими дисциплинами. Компоненты систем Data Mining. Основные этапы анализа данных. Исторический обзор	3	1-3			
2.	Тема 2. Предварительная обработка данных. Очистка данных. Интеграция и преобразование данных. Сокращение данных.	3	4-6			
3.	Тема 3. Задача классификации с обучением. Классификация с использованием деревьев решений, нейронных сетей. "Наивная" байесовская классификация, байесовские сети. Метод k ближайших соседей.	3	7-9			
4.	Тема 4. Задача поиска ассоциативных правил на примере анализа рыночной корзины. Метод ?A priori? генерации ассоциативных правил.	3	10-12			

N	Раздел Дисциплины	Семестр	Неделя семестра	Виды самостоятельной работы студентов	Трудоемкость (в часах)	Формы контроля самостоятельной работы
5.	Тема 5. Кластерный анализ. Типы данных в кластерном анализе. Методы k средних, k медоидов. Агломеративные и дивизимные методы иерархической кластеризации.	3	13-15			
6.	Тема 6. Введение в нечеткую логику. Определение нечетких множеств. Основные операции с нечеткими множествами. Алгоритм не-четкой кластеризации.	3	16-18			
	Итого				0	

5. Образовательные технологии, включая интерактивные формы обучения

Обучение происходит в форме лабораторных занятий и самостоятельной работы студентов. В течение семестра в рамках текущей работы каждый студент разрабатывает презентацию на выбранную тему и делает доклад на семинаре. Проведение доклада сопровождается и завершается обсуждением изучаемой темы.

6. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины и учебно-методическое обеспечение самостоятельной работы студентов

Тема 1. Введение в интеллектуальный анализ данных. Области использования. Связь с другими дисциплинами. Компоненты систем Data Mining. Основные этапы анализа данных. Исторический обзор

Тема 2. Предварительная обработка данных. Очистка данных. Интеграция и преобразование данных. Сокращение данных.

Тема 3. Задача классификации с обучением. Классификация с использованием деревьев решений, нейронных сетей. "Наивная" байесовская классификация, байесовские сети. Метод k ближайших соседей.

Тема 4. Задача поиска ассоциативных правил на примере анализа рыночной корзины. Метод ?A priori? генерации ассоциативных правил.

Тема 5. Кластерный анализ. Типы данных в кластерном анализе. Методы k средних, k медоидов. Агломеративные и дивизимные методы иерархической кластеризации.

Тема 6. Введение в нечеткую логику. Определение нечетких множеств. Основные операции с нечеткими множествами. Алгоритм не-четкой кластеризации.

Тема . Итоговая форма контроля

Тема . Итоговая форма контроля

Тема . Итоговая форма контроля

Примерные вопросы к зачету:

По данному курсу предусмотрен зачет. К зачету каждый студент самостоятельно выбирает предметную область и среду программирования, затем формулирует и решает в этой среде программирования одну из задач интеллектуального анализа данных.

7.1. Основная литература:

1. Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. Классификация и снижение размерности. - М.: Финансы и статистика, 1989.
2. Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И. Методы и модели анализа данных: OLAP и Data Mining. - СПб.: БХВ-Петербург, 2004.
3. Кофман А. Введение в теорию нечетких множеств. - М.: Радио и связь, 1982.
4. Степанов Р.Г. Технология Data Mining: Интеллектуальный анализ данных. - Учебное пособие. Казанский государственный университет, 2009.

7.2. Дополнительная литература:

1. Han J., Kamber M., Data mining: Concepts and Techniques. - Morgan Kaufmann Publishers, 2001.
2. Konar A., Artificial intelligence and soft computing: behavioral and cognitive modeling of the human brain. - CRC Press LLC. - Boca Raton, Florida, 2000.
3. Mitra S., Acharya T., Data Mining. Multimedia, Soft Computing, and Bioinformatics. - John Wiley & Sons, Inc. - Hoboken, New Jersey, 2003.

7.3. Интернет-ресурсы:

8. Материально-техническое обеспечение дисциплины/модуля согласно утвержденному учебному плану

Освоение дисциплины "Научно-исследовательский семинар" предполагает использование следующего материально-технического обеспечения:

Программа составлена в соответствии с требованиями ФГОС ВПО и учебным планом по направлению 010400.68 "Прикладная математика и информатика" и магистерской программе Математическое и информационное обеспечение экономической деятельности .

Автор(ы):

"__" _____ 201__ г.

Рецензент(ы):

Миссаров М.Д. _____

"__" _____ 201__ г.