

ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«Приднестровский государственный университет им. Т.Г. Шевченко»

Рыбницкий филиал

Кафедра прикладной информатики в экономике

УТВЕРЖДАЮ

Директор филиала ПГУ
им. Т.Г. Шевченко в г. Рыбница
профессор



Павлинов И.А.

“ 22 ” 09 2018 г.

РАБОЧАЯ ПРОГРАММА

на 2018 / 2019 учебный год

УЧЕБНОЙ ДИСЦИПЛИНЫ

«Модели и методы интеллектуального анализа данных»

Направление подготовки:

09.04.03 «Прикладная информатика»

Профиль подготовки:

«Информационные технологии в моделировании и организации бизнес-процессов»

квалификация (степень) выпускника

Магистр

Форма обучения:

очная

Рыбница, 2018

Рабочая программа дисциплины «*Модели и методы интеллектуального анализа данных*» / сост. В.Н. Черний – Рыбница: ГОУ ВО «ПГУ им. Т.Г. Шевченко», 2018. – 21 с.

**РАБОЧАЯ ПРОГРАММА ПРЕДНАЗНАЧЕНА ДЛЯ ПРЕПОДАВАНИЯ ДИСЦИПЛИНЫ
БАЗОВОЙ ЧАСТИ БЛОКА ДИСЦИПЛИН СТУДЕНТАМ ОЧНОЙ ФОРМЫ ОБУЧЕНИЯ ПО
НАПРАВЛЕНИЮ ПОДГОТОВКИ 09.04.03 «ПРИКЛАДНАЯ ИНФОРМАТИКА»**

Рабочая программа составлена с учетом Федерального Государственного образовательного стандарта высшего образования по направлению подготовки 09.04.03 «Прикладная информатика» (уровень магистратуры), утвержденного приказом №1404 Министерства образования и науки Российской Федерации от 30.10.14 г.

Составитель _____ / Черний Валентина Николаевна, ст. преподаватель/

1. Цели и задачи освоения дисциплины:

Дисциплина «Модели и методы интеллектуального анализа данных» является частью профессионального цикла дисциплин подготовки студентов по направлению 09.03.03 «Прикладная информатика».

Цель курса – рассмотрение вопросов хранения и представления информации, изучение современных методов обработки и анализа данных, в том числе методов и моделей интеллектуального анализа данных (ИАД). Формирование представления о типах задач, возникающих в области интеллектуального анализа данных (Data Mining) и методах их решения, которые помогут обучающимся выявлять, формализовать и успешно решать практические задачи анализа данных, возникающие в процессе их профессиональной деятельности.

Задачи курса:

- изучение методов и моделей Data Mining;
- получение представления об алгоритмах построения деревьев решений;
- изучение алгоритмов классификации и регрессии;
- изучение алгоритмов поиска ассоциативных правил;
- изучение методов кластеризации.

2. Место дисциплины в структуре ООП ВО:

Дисциплина входит в профессиональный цикл дисциплин федерального компонента направления подготовки 09.04.03 «Прикладная информатика», адресована для подготовки магистров по магистерской программе «Информационные технологии в моделировании и организации бизнес-процессов» и обеспечивает теоретическую и практическую подготовку студентов в области архитектуры современных информационных систем. Курс является базовым и на его основе строятся последующие курсы в сфере информационных систем и технологий.

Дисциплине «Модели и методы интеллектуального анализа данных» предшествуют дисциплины: «Математическое моделирование»; «Математические и инструментальные средства поддержки принятия решений»; «Архитектура современных информационных систем».

Дисциплина «Модели и методы интеллектуального анализа данных» предшествует дисциплинам: «Геоинформационные системы и технологии»; «Компьютерные технологии анализа и прогнозирования в экономических системах»; «Управление инновационными проектами».

3. Требования к результатам освоения дисциплины:

Изучение дисциплины направлено на формирование следующих компетенций:

Код компетенции	Формулировка компетенции
<i>Общепрофессиональные компетенции:</i>	
ОПК-4	способностью исследовать закономерности становления и развития информационного общества в конкретной прикладной области
ОПК-6	способностью к профессиональной эксплуатации современного электронного оборудования в соответствии с целями ООП магистратур
<i>Профессиональные компетенции:</i>	
ПК-3	способностью ставить и решать прикладные задачи в условиях неопределенности и определять методы и средства их эффективного

4.2. *Распределение видов учебной работы и их трудоемкости по разделам дисциплины.*

№ раздела	Наименование разделов	Количество часов				
		Всего	Аудиторная работа			Внеауд. работа (СР)
			Л	ПЗ	ЛР	
1.	Хранение информации	24	6	–	6	12
2.	Современные методы обработки и анализа данных	30	8	–	8	14
3	Основные понятия, задачи, стадии и методы интеллектуального анализа данных	24	6	–	6	12
4	Построение и использование моделей для анализа данных	30	8	–	8	14
	<i>Итого:</i>	<i>144</i>	<i>28</i>	<i>–</i>	<i>28</i>	<i>52/36</i>
	<i>Всего:</i>	<i>144</i>	<i>28</i>	<i>–</i>	<i>28</i>	<i>52/36</i>

4.3. *Тематический план по видам учебной деятельности*

Лекции

№ п/п	Номер раздела дисциплины	Объем часов	Тема лекции	Учебно-наглядные пособия
1	1	2	Основные понятия интеллектуального анализа данных (ИАД)	Интерактивная презентация
2	1	2	Средства для поиска моделей и управления объектами ИАД	Интерактивная презентация
3	1	2	Выбор правильного алгоритма в аналитической задаче	Интерактивная презентация
4	2	2	Алгоритмы интеллектуального анализа данных	Интерактивная презентация
5	2	2	Анализ проблемы интеграции данных, знакомство с OLTP и OLAP-технологиями	Интерактивная презентация
6	2	2	Запросы интеллектуального анализа данных	Интерактивная презентация
7	2	2	Определение алгоритма, используемого моделью интеллектуального анализа данных	Интерактивная презентация
8	3	2	История развития OLAP. Архитектуры OLAP.	Интерактивная презентация
9	3	2	Компоненты OLAP-систем. Обзор продуктов OLAP.	Интерактивная презентация
10	3	2	Задача поиска ассоциативных правил. Задача кластеризации.	Интерактивная презентация
11	4	2	Добыча данных – Data Mining.	Интерактивная презентация
12	4	2	Задачи Data Mining.	Интерактивная презентация
13	4	2	Классификация задач Data Mining.	Интерактивная презентация
14	4	2	Задача классификации и регрессии.	Интерактивная презентация
	Итого:	28		

Самостоятельная работа студента

Раздел дисциплины	№ п/п	Тема и вид СРС	Трудоемкость (в часах)
Раздел 1	1	Методы и стадии Data Mining. Сферы применения Data Mining.	12
Раздел 2	2	Сравнительная характеристика OLTP и OLAP.	14
Раздел 3	3	Построение кубов данных. Создание источника данных.	12
Раздел 4	4	Архитектура хранилища HOLAP, ROLAP и MOLAP.	14

5. Примерная тематика курсовых проектов (работ)

Курсовая работа по учебной дисциплине «Модели и методы интеллектуального анализа данных» учебным планом не предусмотрена.

6. Образовательные технологии

Рекомендуемые образовательные технологии: лекции, лабораторные работы, самостоятельная работа студентов.

Необходимо использовать активные и интерактивные формы обучения (разбор конкретных ситуаций, обсуждение отдельных разделов дисциплины).

В сочетании с внеаудиторной работой это способствует формированию и развитию профессиональных навыков обучающихся.

Для закрепления знаний студентов по каждому разделу курса «Модели и методы интеллектуального анализа данных» проводятся практические занятия, целью которых является формирование навыков самостоятельной работы по решению задач интеллектуального анализа данных.

Планирование времени на самостоятельную работу, необходимого на изучение настоящей дисциплины, лучше всего осуществлять на весь семестр, предусматривая при этом регулярное повторение пройденного материала.

Материал, законспектированный на лекциях, необходимо регулярно дополнять сведениями из литературных источников, представленных в списке рекомендуемой литературы. По каждой из тем для самостоятельного изучения, приведенных в программе дисциплины, следует сначала прочитать рекомендованную литературу и при необходимости составить краткий конспект основных положений, терминов, сведений, требующих запоминания и являющихся основополагающими в этой теме и нужных для освоения последующих разделов.

Для расширения знаний по дисциплине рекомендуется использовать Интернет-ресурсы: проводить поиск в различных поисковых системах, таких как www.yandex.ru, www.google.ru, www.yahoo.ru и использовать материалы сайтов, рекомендованных преподавателем на занятиях.

Семестр	Вид занятия (Л, ПР, ЛР)	Используемые интерактивные образовательные технологии	Количество часов
3	Л	Классы с компьютером и мультимедиа проектором	28
3	ЛР	Классы с компьютерами и мультимедиа проектором	28

- с. интеллектуальный.
7. Какой класс задач анализа отвечает за поиск закономерностей в данных?
- информационно-поисковый;
 - оперативно-аналитический;
 - интеллектуальный.
8. Какая подсистема СППР отвечает за ввод данных?
- OLTP;
 - хранилище данных;
 - SQL;
 - OLAP;
 - Data Mining.
9. Какая подсистема СППР отвечает за хранение данных?
- OLTP;
 - хранилище данных;
 - SQL;
 - OLAP;
 - Data Mining.
10. Какая подсистема СППР отвечает за информационно-поисковый анализ данных?
- OLTP;
 - хранилище данных;
 - SQL;
 - OLAP;
 - Data Mining.
11. Какая подсистема СППР отвечает за оперативный анализ данных?
- OLTP;
 - хранилище данных;
 - SQL;
 - OLAP;
 - Data Mining.
12. Какая подсистема СППР отвечает за интеллектуальный анализ данных?
- OLTP;
 - хранилище данных;
 - SQL;
 - OLAP;
 - Data Mining.
13. Как реализуется подсистема ввода данных?
- с помощью технологии Data Mining;
 - с помощью базы данных;
 - с помощью СУБД;
 - с помощью хранилища данных;
 - с помощью витрины данных.
14. Какие данные могут храниться в системе анализа?
- детализированные;
 - обобщенные;
 - детализированные и обобщенные.

8. Как называются числовые фактические данные, которые могут быть просуммированы по всем измерениям?
 - a. аддитивные;
 - b. полуаддитивные;
 - c. неаддитивные.

9. Как называются числовые фактические данные, которые могут быть просуммированы только по определенным измерениям?
 - a. аддитивные;
 - b. полуаддитивные;
 - c. неаддитивные.

10. Как называются фактические данные, которые не могут быть просуммированы ни по одному измерению?
 - a. аддитивные;
 - b. полуаддитивные;
 - c. неаддитивные.

11. На какие вопросы должны отвечать метаданные?
 - a. что, кто, где, как, когда, почему;
 - b. что, кто, зачем, как, когда, почему;
 - c. что, кто, где, как, по какой причине, почему;
 - d. что, кто, где, как, когда, сколько.

12. Репозиторий – это
 - a. словарь терминов;
 - b. хранилище метаданных;
 - c. каталог с файлами.

13. Если поток образуется данными, копируемыми из ОИД, то он называется
 - a. входным потоком;
 - b. потоком обобщения;
 - c. архивным потоком;
 - d. потоком метаданных;
 - e. обратным потоком.

14. Если поток образуется агрегированными данными, то он называется
 - a. входным потоком;
 - b. потоком обобщения;
 - c. архивным потоком;
 - d. потоком метаданных;
 - e. обратным потоком.

15. Если поток образуется детальными данными, количество обращений к которым снизилось, то он называется
 - a. входным потоком;
 - b. потоком обобщения;
 - c. архивным потоком;
 - d. потоком метаданных;
 - e. обратным потоком.

16. Если поток образуется данными, переносимыми в репозиторий, то он называется

- a. MOLAP;
- b. ROLAP;
- c. HOLAP;
- d. DOLAP;
- e. JOLAP.

7. Настольная OLAP, предназначенная для локального анализа и представления данных, называется

- a. MOLAP;
- b. ROLAP;
- c. HOLAP;
- d. DOLAP;
- e. JOLAP.

8. OLAP, предназначенная для создания и управления данными и метаданными, называется

- a. MOLAP;
- b. ROLAP;
- c. HOLAP;
- d. DOLAP;
- e. JOLAP.

9. В каком отношении находятся таблица фактов и таблица измерений?

- a. «один-к-одному»;
- b. «один-ко-многим»;
- c. «многие-ко-многим».

10. Исследование и обнаружение машиной (алгоритмами, средствами искусственного интеллекта) в сырых данных скрытых знаний, которые: ранее не были известны, нетривиальны, практически полезны, доступны для интерпретации человеком, называется

- a. OLTP;
- b. хранилищем данных;
- c. OLAP;
- d. Data Mining.

11. Какие операции над данными включены в ETL-процесс?

- a. ввод, модификация, вывод;
- b. чтение, изменение, запись;
- c. извлечение, преобразование, загрузка;
- d. получение, хранение, анализ.

12. Гиперкуб – это ...

- a. объект, все измерения которого имеют одинаковую размерность;
- b. поликуб;
- c. объект, все измерения которого имеют разную размерность;
- d. многомерный куб;
- e. многомерная база данных.

13. Многомерный просмотр данных основан на ...

- a. многомерной базе данных;
- b. технологии мультимедиа;
- c. многослойной базе;
- d. сетевой технологии;

6. Задача регрессии сводится к ...
- нахождения частых зависимостей между объектами или событиями;
 - определения класса объекта по его характеристиками;
 - определение по известным характеристиками объекта значение некоторого его параметра;
 - поиска независимых групп и их характеристик в всем множестве анализируемых данных.
7. Задача кластеризации заключается в ...
- нахождения частых зависимостей между объектами или событиями;
 - определения класса объекта по его характеристиками;
 - определение по известным характеристиками объекта значение некоторого его параметра;
 - поиска независимых групп и их характеристик в всем множестве анализируемых данных.
8. Целью поиска ассоциативных правил является ...
- нахождения частых зависимостей между объектами или событиями;
 - определения класса объекта по его характеристиками;
 - определение по известным характеристиками объекта значение некоторого его параметра;
 - поиска независимых групп и их характеристик в всем множестве анализируемых данных.
9. В описательных моделях относятся следующие модели данных:
- модели классификации и последовательностей;
 - регрессивные, кластеризации, исключений, итоговые и ассоциации;
 - классификации, кластеризации, исключений, итоговые и ассоциации;
 - модели классификации, последовательностей и исключений.
10. Модели классификации описывают ...
- правила или набор правил в соответствии с которыми можно отнести описание любого нового объекта к одному из классов;
 - функции, которые позволяют прогнозировать изменения непрерывных числовых параметров;
 - функциональные зависимости между зависимыми и независимыми показателями и переменными в понятной человеку форме;
 - группы, на которые можно разделить объекты, данные о которых подвергаются анализа.
11. Модели последовательностей описывают ...
- правила или набор правил в соответствии с которыми можно отнести описание любого нового объекта к одному из классов;
 - функции, которые позволяют прогнозировать изменения непрерывных числовых параметров;
 - функциональные зависимости между зависимыми и независимыми показателями и переменными в понятной человеку форме;
 - группы, на которые можно разделить объекты, данные о которых подвергаются анализа.
12. Регрессивные модели описывают ...
- правила или набор правил в соответствии с которыми можно отнести описание

8. При слиянии данных в информационное хранилище из внутренних и внешних источников обеспечивается ...
- предметная ориентация данных;
 - выбор требуемых сведений из предметных приложений по наименованиям;
 - гипертекстовый просмотр данных;
 - согласование данных по наименованию;
 - хранение данных по предметным областям.
9. Интеллектуальный выбор данных из информационного хранилища – это ...
- реализация методов искусственного интеллекта;
 - выбор по заданному алгоритму;
 - реализация самообучающихся систем;
 - реализация экономико-статистических методов.
10. В процессе погружения в информационное хранилище данные ...
- очищаются от ненужной для анализа информации;
 - агрегируются;
 - преобразуются из разных типов данных предметных приложений в единую структуру хранения;
 - индексируются;
 - синхронизируются.
11. Какова основная цель создания и использования хранилищ данных:
- анализ данных для принятия управленческих решений;
 - надежное хранение накопленных данных;
 - резервное копирование данных.

7.2. Перечень вопросов для подготовки к итоговому контролю (экзамену)

- Область применения хранилищ данных.
- Основные требования к данным, вводимым в хранилище данных.
- Возможности создания хранилищ данных в реляционных БД.
- Понятие о витринах данных.
- Понятие многомерного представления при описании структур данных.
- Алгоритм добычи данных – метод деревьев решений.
- Алгоритм добычи данных – кластеризация.
- Технология аналитической обработки данных (OLAP).
- Варианты реализации OLAP.
- Распределенные корпоративные хранилища данных.
- Зависимые и независимые киоски данных.
- Характеристика виртуальных кубов.
- Архитектура хранилища HOLAP.
- Архитектура хранилища ROLAP.
- Архитектура хранилища MOLAP.
- Особенности ETL-процесса.
- Методы и этапы процесса очистки данных.
- Задачи Data Mining. Классификация задач DataMining.
- Задача классификации и регрессии.
- Задача поиска ассоциативных правил.
- Задача кластеризации.
- Модели DataMining.
- Методы DataMining.

2. CITForum (<http://www.citforum.ru>). На сайте представлена огромная библиотека материалов по информационным технологиям. Все они опубликованы на некоммерческой основе.

3. Портал iXBT.com (<http://www.ixbt.com>). Сайт iXBT.com создан и развивается с одной стратегической целью – предоставить возможность получить как можно более полную, объективную и полезную информацию о высоких технологиях, персональных компьютерах, их компонентах и периферийных устройствах.

4. Портал 3DNews (<http://www.3dnews.ru>). 3DNews Daily Digital Digest – первое независимое российское онлайн издание, посвященное компьютерным технологиям. Существовая с 17-го июля 1997 года, проект без каких-либо сторонних инвестиций смог вырасти в успешно развивающуюся интернет-компанию и занять свою уникальную нишу информационного пространства Рунета.

9. Материально-техническое обеспечение дисциплины (модуля):

Материально-техническое обеспечение дисциплины «Модели и методы интеллектуального анализа данных» включает перечень аудиторий с установленным в них оборудованием, в которых проводятся аудиторные занятия:

1. Лекционная аудитория, оборудованная видеопроектором для демонстрации учебных материалов (презентаций и демонстрации учебных фильмов), на 25 посадочных мест (ауд.203):

- Акустическая система (2 колонки)
- Ноутбук HP ProBook 4720s
- Мультимедийный проектор
- Выносной экран.

2. Аудитория для проведения лабораторных занятий на 12 посадочных мест (ауд.206).

10. Методические рекомендации по организации изучения дисциплины:

Рабочая программа составлена с учетом Федерального Государственного образовательного стандарта высшего образования по направлению подготовки 09.04.03 «Прикладная информатика» (уровень магистратуры) и учебного плана по магистерской программе «Информационные технологии в моделировании и организации бизнес-процессов».

Изучение дисциплины проходит в форме лекционных, лабораторных и практических занятий.

Видами текущего контроля по дисциплине «Модели и методы интеллектуального анализа данных» является: прием индивидуальных заданий, проверка подготовленных материалов для семинаров и практических занятий, проверка самостоятельной работы.

Итоговый контроль осуществляется в виде экзамена.

11. Технологическая карта дисциплины

Курс 2 группа РФ17ДР68ПЭ семестр 3

Преподаватель-лектор Черний Валентина Николаевна

Преподаватели, ведущие практические занятия Черний Валентина Николаевна

Кафедра прикладной информатики в экономике

Весовой коэффициент дисциплины в совокупной рейтинговой оценке, рассчитываемой по всем дисциплинам (если введена модульно-рейтинговая система) модульно-рейтинговая система введена

Составитель  / Черний Валентина Николаевна, ст. преподаватель/

Зав. кафедрой  / Павлинов Игорь Алексеевич, профессор/

Согласовано:

1. Зав. выпускающей кафедры ПИЭ  / Павлинов Игорь Алексеевич, профессор/

2. Директор филиала ПГУ им. Т.Г. Шевченко в г. Рыбница

 / Павлинов Игорь Алексеевич, профессор/