

Минобрнауки России
**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ**
«ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «ВГУ»)



УТВЕРЖДАЮ
Заведующий кафедрой
Борисов Дмитрий Николаевич
Кафедра информационных систем

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

21.04.21

Б1.В.05 Системы интеллектуального анализа данных

1. Код и наименование направления подготовки/специальности:

09.04.02 Информационные системы и технологии

2. Профиль подготовки/специализация:

Анализ и синтез информационных систем

3. Квалификация (степень) выпускника:

Магистратура

4. Форма обучения:

Очная

5. Кафедра, отвечающая за реализацию дисциплины:

Кафедра информационных систем

6. Составители программы:

ФИО *Сычев Александр Васильевич*

Ученая степень *кандидат физико-математических наук*

Ученое звание *доцент*

E-mail: *sav@cs.vsu.ru*

Факультет: *компьютерных наук*

Кафедра: *информационных систем*

7. Рекомендована: *НМС ФКН, N5 от 10.03.21*

8. Учебный год:

2022-2023

9. Цели и задачи учебной дисциплины:

Целью изучения дисциплины является практическое знакомство студентов с современными технологиями анализа многомерных данных, включая математические модели, алгоритмы и программные средства, используемые для решения основных задач анализа многомерных данных: классификации, кластеризации и других.

Задачи учебной дисциплины:

- изучение структуры систем поддержки принятия решений (СППР), ее отличий от систем оперативной обработки данных (OLTP), многомерной модели данных OLAP и принципов работы с реализацией данной модели (на примере OLAP Analysis Services);
- изучение концепции Data Mining и основных задач, реализующих эту концепцию: классификацию,

кластеризацию, поиск ассоциативных правил и других;

- изучение структуры процесса Data Mining и связанных с ним стандартов;

- практическое изучение задач Data Mining и методов их решения с помощью программных пакетов RapidMiner, Matlab и других.

10. Место учебной дисциплины в структуре ООП:

Учебная дисциплина относится к части блока Б1, формируемой участниками образовательных отношений

11. Планируемые результаты обучения по дисциплине/модулю (знания, умения, навыки), соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями выпускников) и индикаторами их достижения:

Код и название компетенции	Код и название индикатора компетенции	Знания, умения, навыки
ПКВ-16 Способен определять качество проводимых исследований, обрабатывать, интерпретировать и оформлять результаты проведенных исследований и представлять ре-зультаты профессиональному сообществу	ПКВ-16.1 Умеет обрабатывать данные проводимых исследований с использованием современных методов анализа информации и информационных технологий	Умеет: <ul style="list-style-type: none">• сводить задачу анализа многомерных данных к одной из типовых задач Data Mining;• выбирать релевантные методы и алгоритмы решения задач Data Mining;• использовать необходимые программные средства реализации выбранных методов решения задачи интеллектуального анализа многомерных данных;• оценивать получаемые результаты решения задачи Data Mining и вносить изменения в исходную постановку задачи или алгоритм решения этой задачи.

Код и название компетенции	Код и название индикатора компетенции	Знания, умения, навыки
ПКВ-9 Умеет осуществлять моделирование процессов и объектов, постановку и проведение экспериментов по заданной методике и анализ результатов, осуществлять выбор оптимальных решений	ПКВ-9.1 Знает методы исследования предметной области, математические модели описания предметной области, методы оптимизации прикладных задач, современные методики тестирования ИС, методики описания и моделирования бизнес-процессов, средства моделирования бизнес-процессов	Знает: <ul style="list-style-type: none"> • отличительные особенности многомерных данных, способы их представления, хранения и обработки; • технологии представления многомерных данных OLAP, основные операции OLAP; • основные задачи интеллектуального анализа многомерных данных, их свойства, методы их решения; • назначение и структуру систем поддержки принятия решения (СППР), основные задачи, решаемые СППР; • структуру процесса Data Mining и связанные с ним стандарты.
ПК-8 Способен разрабатывать новые технологии проектирования информационных систем, прогнозировать развитие информационных систем и технологий	ПК-8.1 Знает инструменты и методы моделирования бизнес-процессов, современные подходы и стандарты автоматизации организации, отраслевую документацию, основы реинжиниринга бизнес-процессов организации	Знает: <ul style="list-style-type: none"> • стандарты Data Mining; • структуру процесса Data Mining.

12. Объем дисциплины в зачетных единицах/час:

5/180

Форма промежуточной аттестации:

Экзамен

13. Трудоемкость по видам учебной работы

Вид учебной работы	Семестр 3	Всего
Аудиторные занятия	50	50
Лекционные занятия	16	16
Практические занятия		0

Вид учебной работы	Семестр 3	Всего
Лабораторные занятия	34	34
Самостоятельная работа	94	94
Курсовая работа		0
Промежуточная аттестация	36	36
Часы на контроль	36	36
Всего	180	180

13.1. Содержание дисциплины

п/п	Наименование раздела дисциплины	Содержание раздела дисциплины	Реализация раздела дисциплины с помощью онлайн-курса, ЭУМК
1	Введение в Data Mining.	Основные определения, предметная область, актуальность и приложения.	Онлайн курс на edu.vsu.ru
2	Системы поддержки принятия решений и хранилища данных	Системы поддержки принятия решений (СППР). Классы задач анализа данных в СППР. Обобщенная архитектура СППР. OLTP и СППР - сравнение. Понятие об интеллектуальном анализе данных и Data Mining. Концепция хранилища данных (ХД). Свойства ХД. Архитектура систем ХД. Структура СППР с физическим ХД. Структура СППР с виртуальным ХД. Витрина данных (ВД). Структура СППР с самостоятельными ВД. Архитектура ХД.	Онлайн курс на edu.vsu.ru
3	OLAP-системы	Многомерная модель данных. Основные элементы OLAP. Представление данных в виде гиперкуба. Операции над гиперкубом. Определение OLAP-систем. Двенадцать правил Кодда для OLAP. Дополнительные правила Кодда. Тест FASMI. Архитектура OLAP-систем: MOLAP, ROLAP, HOLAP.	Онлайн курс на edu.vsu.ru

п/п	Наименование раздела дисциплины	Содержание раздела дисциплины	Реализация раздела дисциплины с помощью онлайн-курса, ЭУМК
4	Задачи Data Mining	<p>Правила классификации. Деревья решений. Конструирование дерева решений. Критерий расщепления. Алгоритмы: байесовский, CART, C4.5. Алгоритмы классификации: метод "ближайшего соседа", метод построения математических функций, метод опорных векторов (SVM). Постановка задачи. Методы прогнозирования временных рядов. Методы поиска ассоциативных правил. Алгоритм Apriori и его разновидности. Понятие кластерного анализа. Характеристики кластеров. Методы кластерного анализа: иерархические и неиерархические. Иерархические методы кластеризации. Агломеративные и дивизимные методы. Дендрограммы. Метрики подобия объектов. Проверка качества кластеризации. Достоинства и недостатки иерархических алгоритмов. Алгоритм k-средних. Достоинства и недостатки алгоритма. Проверка качества кластеризации. Этапы кластерного анализа. Сложности и проблемы кластерного анализа. Сравнение иерархических и неиерархических методов кластеризации. Этапы визуального анализа данных. Характеристики средств визуализации данных. Типы методов визуализации. Визуализация Data Mining моделей. Методы визуализации. Параллельные координаты. "Лица Чернова". Рекомендации по использованию средств визуализации. Качество визуализации. Основные тенденции в области визуализации.</p>	Онлайн курс на edu.vsu.ru
5	Стандарты Data Mining	<p>Аспекты стандартизации Data Mining. Стандарты унификации интерфейсов. Стандарт CWM. Стандарт CRISP. Методология CRISP-DM. SEMMA методология. Стандарт PMML.</p>	Онлайн курс на edu.vsu.ru
6	Процесс Data Mining	<p>Этапы процесса Data Mining. Анализ предметной области. Постановка задачи. Подготовка данных. "Грязные" данные. Очистка данных. Этапы очистки данных. Моделирование. Организационные факторы Data Mining. Человеческие факторы. Роли в Data Mining.</p>	Онлайн курс на edu.vsu.ru

13.2. Темы (разделы) дисциплины и виды занятий

№ п/п	Наименование темы (раздела)	Лекционные занятия	Практические занятия	Лабораторные занятия	Самостоятельная работа	Всего
1	Введение в Data Mining	1			5	6
2	Системы поддержки принятия решений и хранилища данных	2			10	12
3	OLAP-системы	2		4	10	16
4	Задачи Data Mining	8		30	60	98
5	Стандарты Data Mining	1			4	5
6	Процесс Data Mining	2			5	7
		16	0	34	94	144

14. Методические указания для обучающихся по освоению дисциплины

Электронный курс на образовательном портале «Электронный университет ВГУ».
-<https://edu.vsu.ru/course/view.php?id=2488>

15. Перечень основной и дополнительной литературы, ресурсов интернет, необходимых для освоения дисциплины

№ п/п	Источник
1	Зайцев, К. С. Применение методов Data Mining для поддержки процессов управления IT-услугами : учебное пособие / К. С. Зайцев. — Москва : НИЯУ МИФИ, 2009. — 96 с. — ISBN 978-5-7262-1150-3. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/75805
2	Макшанов, А. В. Большие данные. Big Data : учебник для вузов / А. В. Макшанов, А. Е. Журавлев, Л. Н. Тындыкарь. — Санкт-Петербург : Лань, 2021. — 188 с. — ISBN 978-5-8114-6810-2. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/165835

б) дополнительная литература:

№ п/п	Источник
1	Юре, Л. . Анализ больших наборов данных [Электронный ресурс] / Юре Л. , Ананд Р. , Джеффри Д. У. — Москва : ДМК Пресс, 2016 .— 498 с. — <URL:https://e.lanbook.com/book/93571>
2	Чубукова, И. А. Data Mining : учебное пособие / И. А. Чубукова. — 2-е изд. — Москва : ИНТУИТ, 2016. — 470 с. — ISBN 978-5-94774-819-2. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/100582

в) информационные электронно-образовательные ресурсы:

№ п/п	Источник
1	Чубукова И.А. Data Mining (Электронный курс) / Интернет-Университет Информационных Технологий.2006. - [Электрон. ресурс]. Режим доступа: http://www.intuit.ru/department/database/datamining/
2	Электронный курс на образовательном портале «Электронный университет ВГУ». - https://edu.vsu.ru/course/view.php?id=2488

16. Перечень учебно-методического обеспечения для самостоятельной работы

№ п/п	Источник
1	Электронный курс на образовательном портале «Электронный университет ВГУ». - https://edu.vsu.ru/course/view.php?id=2488

17. Образовательные технологии, используемые при реализации учебной дисциплины, включая дистанционные образовательные технологии (ДОТ), электронное обучение (ЭО), смешанное обучение):

Программные пакеты: RapidMiner Studio, Matlab

18. Материально-техническое обеспечение дисциплины:

1. Лекционная аудитория, оборудованная мультимедийным проектором.
2. Компьютерный класс факультета для проведения лабораторных занятий.

19. Оценочные средства для проведения текущей и промежуточной аттестаций

Порядок оценки освоения обучающимися учебного материала определяется содержанием следующих разделов дисциплины:

№ п/п	Разделы дисциплины (модули)	Код компетенции	Код индикатора	Оценочные средства для текущей аттестации
1	3-4	ПКВ-16	ПКВ-16.1	1. КИМы 2. Практические задания

№ п/п	Разделы дисциплины (модули)	Код компетенции	Код индикатора	Оценочные средства для текущей аттестации
2	1-6	ПКВ-9	ПКВ-9.1	1. КИМы 2. Практические задания
3	5-6	ПКВ-8	ПКВ-8.1	1. КИМы

Промежуточная аттестация

Форма контроля - Экзамен

Оценочные средства для промежуточной аттестации

- Оценка «отлично» выставляется в том случае, если студентом полностью выполнено практическое задание, подготовлен отчет и на экзамене студент дал полный ответ на оба вопроса в КИМе;
- оценка «хорошо» выставляется в том случае, если студентом полностью выполнено практическое задание, подготовлен отчет и на экзамене студент дал полный ответ на один вопрос в КИМе и неполный ответ на второй вопрос;
- оценка «удовлетворительно» выставляется в том случае, если студентом выполнено практическое задание, подготовлен отчет и на экзамене студент дал частичный ответ на оба вопроса в КИМе;
- оценка «неудовлетворительно» выставляется в том случае, если студентом не выполнено практическое задание или на экзамене студент дал неправильный ответ на вопросы в КИМе.

20 Типовые оценочные средства и методические материалы, определяющие процедуры оценивания

20.1 Текущий контроль успеваемости

МИНОБРНАУКИ РОССИИ

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ**

**«ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «ВГУ»)**

Кафедра информационных систем

Комплект заданий

по дисциплине Системы интеллектуального анализа данных

Задание. Изучение модели классификации Decision Tree (дерево принятия решения).

В пакете RapidMiner загрузите процесс, демонстрирующий работу модели Decision Tree.

Необходимо исследовать работу данной модели, соединив надлежащим образом входы и выходы блоков процесса.

Итогом исследования должен быть отчет, содержащий:

- Скриншот схемы процесса, содержащий основные блоки и их соединения.
- Описание исходного табличного набора данных для анализа - закладка *DecisionTree.exampleSet(exampleSet)* -> *DataView*
- Скриншот *Tree (Decision Tree)* - *Graph View* и *Text View*.

Необходимо изучить и представить результаты анализа при различных параметрах (выбираются вручную в окне схемы процесса) блока *Decision Tree*.

20.2 Промежуточная аттестация

УТВЕРЖДАЮ

заведующий кафедрой информационных систем.

____.____.20__

Направление подготовки / специальность 09.04.02 "Информационные системы и технологии"

Дисциплина Системы интеллектуального анализа данных.

Вид контроля экзамен.

Контрольно-измерительный материал № 1

1. Хранилища данных (ХД). Свойства ХД. Архитектура систем ХД.
2. Этапы процесса Data Mining .

Преподаватель _____ Сычев А.В.