

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

УТВЕРЖДАЮ

Заведующий кафедрой

Борисов Дмитрий Николаевич

Кафедра информационных систем

28.02.2022



РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

Б1.В.07 Интеллектуальный анализ данных

1. Код и наименование направления подготовки/специальности:

02.04.01 Математика и компьютерные науки

2. Профиль подготовки/специализация:

Компьютерное моделирование и искусственный интеллект

3. Квалификация (степень) выпускника:

Магистратура

4. Форма обучения:

Очная

5. Кафедра, отвечающая за реализацию дисциплины:

Кафедра информационных систем

6. Составители программы:

Сычев Александр Васильевич, кандидат физико-математических наук, доцент кафедры информационных систем

7. Рекомендована: *протокол НМС ФКН №3 от 25.02.2022*

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, курс по выбору

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
ПК-1 Способен демонстрировать фундаментальные знания математических и естественных наук, программирования и информационных технологий.	ПК-1.1 Обладает фундаментальными знаниями, полученными в области математических и (или) естественных наук, программирования и информационных технологий	Знает основные задачи, решаемые в рамках направления информационных технологий Data Mining, их характеристики и основные алгоритмы их решения
ПК-8 Способен создавать и исследовать новые математические модели в естественных науках, промышленности и бизнесе, с учетом возможностей современных информационных технологий, программирования и компьютерной техники	ПК-8.1 Знает основные методы проектирования и производства программного продукта, принципы построения, структуры и приемы работы с инструментальными средствами, поддерживающими создание программных продуктов и программных комплексов, их сопровождения, администрирования и развития (эволюции)	Знает структуру процесса Data Mining, ключевые модули и операторы для конструирования процесса анализа данных в среде RapidMiner

Код и название компетенции	Код и название индикатора компетенции	Знания, умения, навыки
<p>ПК-9 Способен использовать современные методы разработки и реализации конкретных алгоритмов математических моделей на базе языков программирования и пакетов прикладных программ моделирования.</p>	<p>ПК-9.1 Владеет современными методами разработки и реализации алгоритмов математических моделей на базе языков и пакетов прикладных программ моделирования</p>	<p>Имеет навыки реализации основных алгоритмов для различных моделей классификации в среде RapidMiner</p>
<p>ПК-9 Способен использовать современные методы разработки и реализации конкретных алгоритмов математических моделей на базе языков программирования и пакетов прикладных программ моделирования.</p>	<p>ПК-9.2 Умеет разрабатывать и реализовывать алгоритмы математических моделей на базе языков и пакетов прикладных программ моделирования</p>	<p>Умеет реализовывать основные алгоритмы для различных моделей классификации в среде RapidMiner</p>
<p>ПК-9 Способен использовать современные методы разработки и реализации конкретных алгоритмов математических моделей на базе языков программирования и пакетов прикладных программ моделирования.</p>	<p>ПК-9.3 Имеет практический опыт разработки и реализации алгоритмов на базе языков и пакетов прикладных программ моделирования</p>	<p>Имеет опыт реализации основных алгоритмов для различных моделей классификации в среде RapidMiner на примере нескольких типовых наборов данных</p>

Код и название компетенции	Код и название индикатора компетенции	Знания, умения, навыки
<p>ПК-8 Способен создавать и исследовать новые математические модели в естественных науках, промышленности и бизнесе, с учетом возможностей современных информационных технологий, программирования и компьютерной техники</p>	<p>ПК-8.2 Умеет использовать методы проектирования и производства программного продукта, принципы построения, структуры и приемы работы с инструментальными средствами, поддерживающими создание программного продукта</p>	<p>Умеет конструировать процесс анализа данных для решения типовых задач Data Mining в среде RapidMiner, а также настраивать параметры модулей, из которых состоит процесс</p>
<p>ПК-8 Способен создавать и исследовать новые математические модели в естественных науках, промышленности и бизнесе, с учетом возможностей современных информационных технологий, программирования и компьютерной техники</p>	<p>ПК-8.3 Имеет практический опыт применения указанных выше методов и технологий</p>	<p>Имеет опыт конструирования процесса анализа данных для решения задач Data Mining в среде RapidMiner, а также настройки параметров модулей, из которых состоит процесс, на примере типовых наборов данных</p>
<p>ПК-1 Способен демонстрировать фундаментальные знания математических и естественных наук, программирования и информационных технологий.</p>	<p>ПК-1.2 Умеет находить, формулировать и решать стандартные задачи в собственной научно-исследовательской деятельности в области программирования и информационных технологий</p>	<p>Умеет решать стандартные задачи анализа многомерных данных на основе технологий Data Mining</p>

Код и название компетенции	Код и название индикатора компетенции	Знания, умения, навыки
ПК-1 Способен демонстрировать фундаментальные знания математических и естественных наук, программирования и информационных технологий.	ПК-1.3 Имеет практический опыт научно-исследовательской деятельности в области программирования и информационных технологий	Имеет опыт решения стандартных задач анализа многомерных данных на основе технологий Data Mining на примере типовых наборов данных

12. Объем дисциплины в зачетных единицах/час:

3/108

Форма промежуточной аттестации:

Экзамен

13. Трудоемкость по видам учебной работы

Вид учебной работы	Семестр 1	Всего
Аудиторные занятия	36	36
Лекционные занятия	18	18
Практические занятия		0
Лабораторные занятия	18	18
Самостоятельная работа	36	36
Курсовая работа		0
Промежуточная аттестация	36	36
Часы на контроль	36	36
Всего	108	108

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

п/п	Наименование раздела дисциплины	Содержание раздела дисциплины	Реализация раздела дисциплины с помощью онлайн-курса, ЭУМК
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			3	4
2	Системы поддержки принятия решений и хранилища данных	2			6	8
3	OLAP-системы	3		4	8	15
4	Задачи Data Mining	8		14	45	67
5	Стандарты Data Mining	1			5	6
6	Процесс Data Mining	3			5	8
		18	0	18	72	108

14. Методические указания для обучающихся по освоению дисциплины

1) При изучении дисциплины рекомендуется использовать следующие средства:

- рекомендуемую основную и дополнительную литературу;
- методические указания и пособия;
- контрольные задания для закрепления теоретического материала;
- электронные версии учебников и методических указаний для выполнения лабораторно-практических работ.

2) Для лучшего усвоения дисциплины рекомендуется проведение письменного опроса (тестирование, решение задач) студентов по материалам лекций. Подборка вопросов для тестирования осуществляется на основе изученного теоретического материала.

3) При проведении лабораторных занятий обеспечивается практическая демонстрация материалов лекционных занятий и осуществляется экспериментальная проверка методов, алгоритмов и технологий анализа многомерных данных, излагаемых в рамках лекций.

4) При переходе на дистанционный режим обучения для создания электронных курсов, чтения лекций онлайн и проведения лабораторно-практических занятий используются информационные ресурсы образовательного портала "Электронный университет ВГУ (<https://edu.vsu.ru/course/view.php?id=2488>), базирующегося на системе дистанционного обучения Moodle, развернутой в университете.

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	Онлайн курс 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, 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	1-4, 6	ПК-1	ПК-1.1	КИМы
2	4	ПК-8	ПК-8.1	КИМы, Практические задания
3	4,5	ПК-9	ПК-9.1	Практические задания
4	4	ПК-9	ПК-9.2	Практические задания
5	4	ПК-9	ПК-9.3	Практические задания
6	4	ПК-8	ПК-8.2	Практические задания
7	4	ПК-8	ПК-8.3	Практические задания
8	2-4	ПК-1	ПК-1.2	КИМы, Практические задания
9	3,4	ПК-1	ПК-1.3	Практические задания

Промежуточная аттестация

Форма контроля - Экзамен

Оценочные средства для промежуточной аттестации

- оценка «зачтено» выставляется в том случае, если студентом выполнено итоговое практическое задание, подготовлен отчет и при устном собеседовании студент дал частичный ответ на оба вопроса в КИМе;
- оценка «незачтено» выставляется в том случае, если студентом не выполнено итоговое практическое задание или при устном собеседовании 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 Интеллектуальный анализ 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 (выбравтся вручную в окне схемы процесса) блока *Decision Tree*.

20.2 Промежуточная аттестация

УТВЕРЖДАЮ

заведующий кафедрой информационных систем.

__._.20__

Направление подготовки / специальность 02,04,01 "Математика и компьютерные науки".

Дисциплина Интеллектуальный анализ данных.

Вид контроля зачет.

Контрольно-измерительный материал № 1

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

Преподаватель _____ Сычев А.В.