

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.ДВ.03.01 Data mining в бизнес-журналистике

1. Код и наименование специальности:

42.04.02 Журналистика

2. Профиль подготовки/специализация:

Бизнес-журналистика и корпоративные медиакоммуникации

3. Квалификация (степень) выпускника:

Магистратура

4. Форма обучения:

заочная

5. Кафедра, отвечающая за реализацию дисциплины:

Кафедра информационных систем

6. Составители программы: Сычев Александр Васильевич, кандидат физико-математических наук, доцент

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

8. Учебный год:

2023-2024

9. Цели и задачи учебной дисциплины

Целями освоения учебной дисциплины являются:

практическое знакомство студентов с современными технологиями анализа многомерных данных, включая 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), многомерной модели данных OLAP;
- 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 Rapid Miner и других.

10. Место учебной дисциплины в структуре ОПОП:

Учебная дисциплина относится к части, формируемой участниками образовательных отношений, блока Б1 Дисциплины (модули) учебного плана по направлению магистратуры 42.04.02 Журналистика (дисциплины по выбору).

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

Код	Название компетенции	Код(ы)	Индикатор(ы)	Планируемые результаты обучения
УК-1	Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, выработать стратегию действий	УК-1.1	Определяет пробелы в информации, необходимой для решения проблемной ситуации	Знать: принципы системного подхода Уметь: определять пробелы в информации, необходимой для решения проблемной ситуации. Владеть: навыками критического анализа проблемных ситуаций на основе системного подхода.
ПК-1	Способен осуществлять авторскую деятельность любого характера и уровня сложности с учетом специфики разных типов СМИ и других медиа	ПК-1.3.	Анализирует релевантную информацию из доступных документальных источников	Знать: доступные документальные источники Уметь: анализировать релевантную информацию из доступных документальных источников.
		ПК-1.4.	Контролирует достоверность и полноту полученной информации, систематизирует факты и мнения	Уметь: контролировать достоверность и полноту полученной информации, систематизировать факты и мнения. Владеть: приемами и техникой систематизации фактов и мнений.

12. Объем дисциплины в зачетных единицах/час.:
3/108.

Форма промежуточной аттестации:
экзамен

13. Трудоемкость по видам учебной работы

Вид учебной работы	Трудоемкость		
	Всего	По семестрам	
		Курс 2 сессия 2	Курс 2 Сессия 3
Аудиторные занятия	14	2	12
в том числе:	лекции	2	4
	практические	8	8
Самостоятельная работа	90	34	56
Форма промежуточной аттестации – Зачет с оценкой	4		4
Всего:	108	36	72

13.1. Содержание дисциплины

п/п	Наименование раздела дисциплины	Содержание раздела дисциплины	Реализация раздела дисциплины с помощью онлайн-курса, ЭУМК *
1. Лекционные занятия			
1.1	Введение в DataMining	Основные определения, предметная область, актуальность и приложения.	Онлайн курс на edu.vsu.ru
1.2	Стандарты DataMining	Аспекты стандартизации DataMining. Стандарты унификации интерфейсов. Стандарт CWM. Стандарт CRISP. Методология CRISP-DM. SEMMA методология. Стандарт PMML	Онлайн курс на edu.vsu.ru
1.3	Процесс DataMining	Этапы процесса DataMining. Анализ предметной области. Постановка задачи. Подготовка данных. "Грязные" данные. Очистка данных. Этапы очистки данных. Моделирование. Организационные факторы DataMining. Человеческие факторы. Роли в DataMining.	Онлайн курс на edu.vsu.ru
2. Практические занятия			
2.1	Системы поддержки принятия решений и хранилища данных	Системы поддержки принятия решений (СППР). Классы задач анализа данных в СППР. Обобщенная архитектура СППР.	Онлайн курс на edu.vsu.ru

		OLTP и СППР - сравнение. Понятие об интеллектуальном анализе данных и DataMining. 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.	
2.2	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
2.3	Задачи DataMining	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 DataMining моделей. Методы визуализации. Параллельные координаты. "Лица Чернова". Рекомендации по использованию средств визуализации. Качество визуализации. Основные тенденции в области визуализации.	Онлайн курс на edu.vsu.ru

13.2. Темы (разделы) дисциплины и виды занятий

№ п/п	Наименование темы (раздела) дисциплины	Виды занятий (количество часов)			
		Лекции	Практические	Самостоятельная работа	Всего
1	Введение в DataMining	2	-	20	22
2	Системы поддержки принятия решений и хранилища данных	-	2	10	12
3	OLAP-системы	-	2	10	12
4	Задачи DataMining	-	4	10	14
5	Стандарты DataMining	2	-	20	22
6	Процесс DataMining	2	-	20	22
		6	8	90	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 и практических работ. Подборка вопросов для тестирования осуществляется на основе изученного теоретического материала.

Такой подход позволяет повысить мотивацию студентов при конспектировании лекционного материала.

3) При проведении лабораторных занятий обеспечивается практическая демонстрация материалов лекционных занятий и осуществляется экспериментальная проверка методов, алгоритмов и технологий обработки 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 (<http://edu.vsu.ru>), базирующегося на системе дистанционного обучения Moodle, развернутой в университете

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

а) основная литература:

№ п/п	Источник
1	Чубукова, И. А. DataMining : 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
2	Макшанов, А. В. Современные технологии интеллектуального анализа данных : учебное пособие для спо / А. В. Макшанов, А. Е. Журавлев, Л. Н. Тындыкарь. — Санкт-Петербург : Лань, 2020. — 228 с. — ISBN 978-5-8114-5451-8. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/149343
3	Макшанов, А. В. Системы поддержки принятия решений : учебное пособие / А. В. Макшанов, А. Е. Журавлев, Л. Н. Тындыкарь. — Санкт-Петербург :

	Лань, 2020. — 108 с. — ISBN 978-5-8114-5492-1. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/147094
--	--

б) дополнительная литература:

№ п/п	Источник
4	Юре, Л. Анализ больших наборов данных / Л. Юре, Р. Ананд, Д. У. Джеффри ; перевод с английского А. А. Слинкин. — Москва : ДМК Пресс, 2016. — 498 с. — ISBN 978-5-97060-190-7. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/93571

в) информационные электронно-образовательные ресурсы (официальные ресурсы интернет):

№ п/п	Ресурс
1.	Электронная библиотека ЗНБ ВГУ. - https://lib.vsu.ru
2.	ЭБС "Университетская библиотека online". - http://biblioclub.ru/
3.	ЭБС "Консультант студента". - http://www.studmedlib.ru
4.	Ресурсы «Электронного университета». – http://edu.vsu.ru

16. Перечень учебно-методического обеспечения для самостоятельной работы

№ п/п	Источник
1	Электронный курс на образовательном портале «Электронный университет ВГУ». – (https://edu.vsu.ru/)

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

При реализации дисциплины используются различные типы лекций (вводная, обзорная и т. д.), семинарские занятия (проблемные, дискуссионные и т. д.), применяются дистанционные образовательные технологии в части освоения лекционного и практического материала, проведения текущей аттестации, самостоятельной работы по дисциплине или отдельным ее разделам и т. д.

18. Материально-техническое обеспечение дисциплины:

Аудитории для проведения занятий лекционного типа. Типовое оснащение, оборудование: мультимедиапроектор ViewSonic; ПК (i5/4Gb/HDD 1Tb); экран настенный с электроприводом CS 244*244; акустическая система BEHRINGER B115D, микшер UB 1204 FX, микрофон B-1. Программное обеспечение: WinPro 8 RUS Upgrd OLP NL Acdm; OfficeSTd 2013 RUS OLP NL Acdmc; Неисключительные права на ПО Dr. WebEnterpriseSecuritySuite Комплексная защита Dr. WebDesktopSecuritySuite; СПС «ГАРАНТ- Образование».

Аудитории для проведения занятий семинарского типа, текущего контроля и промежуточной аттестации. Типовое оснащение, оборудование: мультимедиапроектор BenQ, экран настенный CS 244*244; переносной ноутбук 15*PackardBell. Программное обеспечение: WinPro 8 RUS Upgrd OLP NL Acdm; OfficeSTd 2013 RUS OLP NL Acdmc; Неисключительные права на ПО Dr. WebEnterpriseSecuritySuite Комплексная защита Dr. WebDesktopSecuritySuite; СПС «ГАРАНТ- Образование».

Аудитории для самостоятельной работы студентов. Используются компьютерные классы: ауд. 115 (Воронеж, ул. Хользунова, 40-а). Типовое оснащение, оборудование: мультимедиапроектор BenQ MX511; экран настенный CS 244*244; интерактивная доска Promethean, ПК (i5/4Gb/HDD 1Tb) (11 шт.);

ауд. 126 (Воронеж, ул. Хользунова, 40-а). Типовое оснащение, оборудование: мультимедиапроектор BenQ MX511; ПК (Razer 5/4Gb/1Tb) (10 шт.); экран настенный CS 244*244, интерактивная доска Promethean.

Программное обеспечение: WinPro 8 RUS Upgrd OLP NL Acdm; OfficeSTd 2013 RUS OLP NL Acdm; Неисключительные права на ПО Dr. WebEnterpriseSecuritySuite Комплексная защита Dr. WebDesktopSecuritySuite; СПС «ГАРАНТ- Образование».

19. Оценочные средства для проведения текущей и промежуточной аттестаций

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

№ п/п	Наименование раздела дисциплины (модуля)	Компетенция(и)	Индикатор(ы) достижения компетенции	Оценочные средства
1.	Введение в DataMining	УК-1	УК-1.1	КИМы
2.	Системы поддержки принятия решений и хранилища данных	ПК-1	ПК-1.3. ПК-1.4.	КИМы
3.	OLAP-системы	ПК-1	ПК-1.3, ПК-1.4	КИМы, практические задания
4.	Задачи DataMining	ПК-1	ПК-1.3, ПК-1.4	КИМы, практические задание
5.	Стандарты DataMining	УК-1	УК-1.1	КИМы
6.	Процесс DataMining	УК-1	УК-1.1	КИМы
Промежуточная аттестация форма контроля: зачет с оценкой				Перечень вопросов Практическое задание

- 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 Data mining в бизнес-журналистике

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 42.04.02 Журналистика.
Дисциплина Data mining в бизнес-журналистике.
Вид контроля Зачет с оценкой.

Контрольно-измерительный материал № 1

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

Преподаватель _____

Сычев А.В.