Подписано электронной подписью:
Вержицкий Данил Григорьевич
Должность: Директор КГПИ КемГУ
Дата и время: 2025-04-23 00:00:00
471086fad29a3b30e244c728abc3661ab35c9d50210dcf0e75e03a5b6fdf6436

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение высшего образования
«КЕМЕРОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
Кузбасский гуманитарно-педагогический институт

Факультет информатики, математики и экономики

УТВЕРЖДАЮ
Декан
А.В. Фомина
«08» февраля 2024 г.

Рабочая программа дисциплины

К.М.09.07 Аналитика данных

Направление подготовки
01.03.02 Прикладная математика и информатика

Направленность (профиль) подготовки
ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ

Программа бакалавриата

Квалификация выпускника
бакалавр

Форма обучения
Очная

Год набора 2024

Новокузнецк 2024

Оглавление

1 Цель дисциплины. 3

Формируемые компетенции, индикаторы достижения компетенций, знания, умения, навыки 3

Место дисциплины....3

2 Объём и трудоёмкость дисциплины по видам учебных занятий. Формы промежуточной аттестации. 3

3. Учебно-тематический план и содержание дисциплины. 4

3.1 Учебно-тематический план 4

4 Порядок оценивания успеваемости и сформированности компетенций обучающегося в текущей и промежуточной аттестации. 4

5 Материально-техническое, программное и учебно-методическое обеспечение дисциплины. 5

5.1 Учебная литература 5

5.2 Материально-техническое и программное обеспечение дисциплины....5

5.3 Современные профессиональные базы данных и информационные справочные системы....6

6 Иные сведения и (или) материалы....6

6.1. Примерные вопросы и задания / задачи для промежуточной аттестации 6

1 Цель дисциплины.

В результате освоения данной дисциплины у обучающегося должны быть сформированы компетенции основной профессиональной образовательной программы бакалавриата (далее - ОПОП): ПК-1

Формируемые компетенции, индикаторы достижения компетенций, знания, умения, навыки

Таблица 1 – Индикаторы достижения компетенций, формируемые дисциплиной

Код и название компетенции	Индикаторы достижения компетенции по ОПОП	Знания, умения, навыки (ЗУВ), формируемые дисциплиной
ПК-1 Способен проводить аналитические исследования с применением технологий больших данных	ПК 1.1 Способен осуществлять выявление, формирование и согласование требований к результатам аналитических работ с применением технологий больших данных ПК 1.2 Способен планировать и организовывать аналитические работы с использованием технологий больших данных ПК 1.4 Способен проводить аналитические исследования с применением технологий больших данных в соответствии с требованиями заказчика	**Знать:** – стандарты обработки и анализа больших данных, и требования, связанные с созданием и использованием систем хранения и обработки больших данных. **Уметь:** – использовать современные инструментальные и вычислительны е средства анализа больших данных, осуществлять постановку задач анализа данных, визуализацию интерпретацию результатов **Владеть:** – навыками обрабатывать и интерпретировать большие данные.

Место дисциплины

Дисциплина включена в модуль «Анализ данных» ОПОП ВО. Дисциплина осваивается на 3 курсе в 6 семестре.

2 Объём и трудоёмкость дисциплины по видам учебных занятий. Формы промежуточной аттестации.

Таблица 2 – Объем и трудоемкость дисциплины по видам учебных занятий

Общая трудоемкость и виды учебной работы по дисциплине, проводимые в разных формах	Объём часов по формам обучения
	ОФО
1 Общая трудоемкость дисциплины	144
2 Контактная работа обучающихся с преподавателем (по видам учебных занятий) (всего)	50
Аудиторная работа (всего):	50
в том числе:	
лекции	16
практические занятия, семинары	34
Внеаудиторная работа (всего):	
3 Самостоятельная работа обучающихся (всего)	94
4 Промежуточная аттестация обучающегося - зачет с оценкой (6 семестр)	

3. Учебно-тематический план и содержание дисциплины.

3.1 Учебно-тематический план

Таблица 3 - Учебно-тематический план очной формы обучения

№ недели п/п	Разделы и темы дисциплины по занятиям	Общая трудоёмкость *(всего час.)*	Трудоемкость занятий (час.) ОФО			Формы текущего контроля и промежуточной аттестации успеваемости
			Аудиторн. занятия		СРС	
			лекц.	лаб.		
Семестр 6						
1	Аналитика больших данных	22	4	6	12	Отчеты по лабораторным работам
2	Когнитивный анализ данных	22	2	6	14	Отчеты по лабораторным работам
3	Классификация данных с использованием детерминированных моделей	24	2	6	16	Отчеты по лабораторным работам
4	Классификация данных на основе статистических моделей	20	2	4	14	Отчеты по лабораторным работам
5	Кластер-анализ	18	2	4	12	Отчеты по лабораторным работам
6	Методы снижения размерностей данных	20	2	4	14	Отчеты по лабораторным работам
7	Прикладные инструменты анализа данных	18	2	4	12	Отчеты по лабораторным работам
ИТОГО		144	16	34	94	

4 Порядок оценивания успеваемости и сформированности компетенций обучающегося в текущей и промежуточной аттестации.

Для положительной оценки по результатам освоения дисциплины обучающемуся необходимо выполнить все установленные виды учебной работы. Оценка результатов работы обучающегося в баллах (по видам) приведена в таблице 4.

Таблица 4 - Балльно-рейтинговая оценка результатов учебной работы обучающихся по видам (БРС)

Учебная работа (виды)	Сумма баллов	Виды и результаты учебной работы	Оценка в аттестации	Баллы
Текущая учебная работа в семестре (Посещение занятий по расписанию и выполнение заданий)	**80**	Лабораторные работы (отчет о выполнении лабораторной работы), 16 работ	**2,6 балла** (пороговое значение) **5 баллов** (максимальное значение)	41-80
Итого по текущей работе в семестре				41-80
Промежуточная аттестация (зачет с оценкой)	20	Теоретический вопрос 1	**2 балла** (пороговое значение) **5 баллов** (максимальное значение)	2 - 5
		Теоретический вопрос 2	**2 балла** (пороговое значение) **5 баллов** (максимальное значение)	2 - 5
		Решение задачи 1.	**3 балла** (пороговое значение) **5 баллов** (максимальное значение)	3 - 5
		Решение задачи 2.	**3 балла** (пороговое значение) **5 баллов** (максимальное значение)	3 - 5

Учебная работа (виды)	Сумма баллов	Виды и результаты учебной работы	Оценка в аттестации	Баллы
Итого по промежуточной аттестации (зачет с оценкой)				10-20
Суммарная оценка по дисциплине:	Сумма баллов текущей и промежуточной аттестации			51 – 100 б.

В промежуточной аттестации оценка выставляется в ведомость в 100-балльной шкале и в буквенном эквиваленте (таблица 5)

Таблица 5 – Соотнесение 100-балльной шкалы и буквенного эквивалента оценки

Сумма набранных баллов	Уровни освоения дисциплины и компетенций	*Экзамен*		*Зачет*
		Оценка	*Буквенный эквивалент*	*Буквенный эквивалент*
86 - 100	Продвинутый	5	отлично	Зачтено
66 - 85	Повышенный	4	хорошо	
51 - 65	Пороговый	3	удовлетворительно	
0 - 50	Первый	2	неудовлетворительно	Не зачтено

5 Материально-техническое, программное и учебно-методическое обеспечение дисциплины.

5.1 Учебная литература

Основная учебная литература

1. Кулаичев, А.П. Методы и средства комплексного анализа данных [Электронный ресурс]: учебн. пособие / А.П. Кулаичев. – Электрон.текстовые дан. – Москва: ИНФРА-М, 2016. – Режим доступа: https://new.znanium.com/read?id=252560

Дополнительная учебная литература

1. Анализ данных : учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2023. — 490 с. — (Высшее образование). — ISBN 978-5-534-00616-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/511020.
2. Миркин, Б. Г. Введение в анализ данных : учебник и практикум / Б. Г. Миркин. — Москва : Издательство Юрайт, 2023. — 174 с. — (Высшее образование). — ISBN 978-5-9916-5009-0. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/511121.

5.2 Материально-техническое и программное обеспечение дисциплины.

Учебные занятия по дисциплине проводятся в учебных аудиториях КГПИ ФГБОУ ВО КемГУ:

610 Учебная аудитория (мультимедийная)для проведения: - занятий лекционного типа; - занятий семинарского (практического) типа; - текущего контроля и промежуточной аттестации. Специализированная (учебная) мебель: доска меловая, кафедра, столы, стулья. Оборудование для презентации учебного материала: стационарное - компьютер, экран, проектор. Используемое программное обеспечение: MS Windows (Microsoft Imagine Premium 3 year по сублицензионному договору № 1212/КМР от 12.12.2018 г. до 12.12.2021 г.), LibreOffice (свободно распространяемое ПО), FoxitReader (свободно распространяемое ПО), Firefox 14 (свободно распространяемое ПО), Яндекс.Браузер (отечественное свободно распространяемое ПО). Интернет с обеспечением доступа в ЭИОС.	654079, Кемеровская область, г. Новокузнецк, пр-кт Металлургов, д. 19
502 Компьютерный класс. Учебная аудитория (мультимедийная) для	654079, Кемеровская область,

проведения: - занятий лекционного типа; - занятий семинарского (практического) типа; - групповых и индивидуальных консультаций; - самостоятельной работы; - текущего контроля и промежуточной аттестации. Специализированная (учебная) мебель: доска меловая, столы компьютерные, стулья. Оборудование для презентации учебного материала: стационарное - компьютер, экран, проектор, наушники. Оборудование: стационарное – компьютеры для обучающихся (16 шт.). Используемое программное обеспечение: MS Windows (Microsoft Imagine Premium 3 year по сублицензионному договору № 1212/КМР от 12.12.2018 г. до 12.12.2021 г.), LibreOffice (свободно распространяемое ПО), FoxitReader (свободно распространяемое ПО), Firefox 14 (свободно распространяемое ПО), Яндекс.Браузер (отечественное свободно распространяемое ПО), Среда статистических вычислений R v.4.0.2 (свободно распространяемое ПО). Интернет с обеспечением доступа в ЭИОС.	г. Новокузнецк, пр-кт Металлургов, д. 19

5.3 Современные профессиональные базы данных и информационные справочные системы.

Перечень СПБД и ИСС по дисциплине

1. Общероссийский математический портал (информационная система) - http://www.mathnet.ru/
2. Научная электронная библиотека eLIBRARY.RU – крупнейший российский информационный портал в области науки, технологии, медицины и образования, содержащий рефераты и полные тексты - www.elibrary.ru
3. База данных Science Direct (более 1500 журналов издательства Elsevier, среди них издания по математике и информатике), режим доступа :https://www.sciencedirect.com

6 Иные сведения и (или) материалы.

6.1. Примерные вопросы и задания / задачи для промежуточной аттестации

1. Аналитика больших данных. Процесс аналитики.
2. Проблема обработки данных. Матрица данных. Гипотезы компактности и скрытых факторов. Структура матрицы данных и задачи обработки. Матрица объект-объект и признак-признак. Расстояние и близость. Измерение признаков. Отношения и их представление.
3. Основные проблемы измерений. Основные типы шкал. Проблема адекватности. Основные задачи анализа и интерпретации данных
4. Стандарты жизненного цикла Big Data: CRISP-DM. Принципы и инструменты аналитики. Задачи и компетенции аналитиков Big Data.
5. Big Data как рынок. Стек технологий аналитика больших данных. Программное обеспечение поддержки принятия решений. Игроки на рынке Big Data. Крупнейшие проекты Big Data в России.
6. Применение математической статистики для анализа данных. Свойства описательных статистик (дескриптивный анализ). Меры изменчивости.
7. Когнитивный анализ данных. Введение в Data Mining – понятие, структура, составляющие и сопутствующие науки. Задачи Data Mining и способы их решения. Классификация методов Data Mining. Области применения Data Mining. Классы систем Data Mining.
8. Процесс накопления и анализа данных: Азбука когнитивного анализа.

9. Данные & знания. Типовые задачи Data Mining. Обучаемые и необучаемые задачи. Жизненный цикл проекта Data Mining. Математический аппарат Data Mining. Стандарты Data Mining.

10. Решающие поверхности и дискриминантные функции. Линейные дискриминантные функции. Линейная разделимость. Кусочно-линейные дискриминантные функции. Нелинейные дискриминантные функции. Фи машины. Потенциальные функции как дискриминантные функции.

11. Пространство весов. Процедуры обучения с коррекцией ошибок: правило с фиксированным приращением, правило абсолютной коррекции, частично корректирующее правило. Обобщенные градиентные методы. Персептронный критерий. Процедуры обучения на основе минимальной среднеквадратичной ошибки: псевдоинверсный метод, метод Хо-Кашпа.

12. Функция потерь. Байесовская дискриминантная функция. Принятие решение по максимуму правдоподобия. Оптимальная дискриминантная функция для нормально распределенных образов. Дискриминантная функция Фишера. Множественный дискриминантный анализ. Пошаговый дискриминантный анализ. Ошибки классификации. Примеры построения статистических дискриминантных функций для различных статистических моделей данных. Обучение для статистических дискриминантных функций.

13. Оценки максимального правдоподобия, байесовские оценки. Непараметрическое оценивание. Парзеновские окна, метод непараметрического оценивания на основе К-ближайшего соседства.

14. Основные типы задач кластер-анализа. Меры подобия и функции расстояния . Выбор критерия кластеризации. Кластерные методы, основанные на евклидовой метрике. Иерархическая кластеризация. Метод К внутригрупповых средних. Использование методов теории графов в задачах кластеризации. Кластеризация на основе анализа плотностей вероятностей.

15. Анализ матриц исходных данных. Метод главных компонент. Корреляционная матрица и ее основные свойства. Собственные векторы и собственные числа корреляционной матрицы. Приведение корреляционной матрицы к диагональной форме. Геометрическая интерпретация главных компонент на плоскости. Модели факторного анализа. Оценка факторных нагрузок методом максимального правдоподобия и центроидным методом.

16. Вращение факторов и их интерпретация. Использование кластеризации признаков для снижения размерности. Многомерное шкалирование.

17. Метрический и неметрический подход к многомерному шкалированию.

18. Методы ортогонального проектирования. Нелинейные методы многомерного шкалирования. Многомерное шкалирование неметрических данных.

19. Многомерные развертки.

20. Готовые комплексные решения: Weka,RapidMiner, Knime, Orange IBM SPSS Modeler (в прошлом Clementine). Инструменты визуализации: Tableau. Фреймворки на JS, D3

Практические задания

Подобрать данные для таблицы, приведенной ниже и проанализировать их взаимное влияние, отобразить корреляцию:

a. Роста ВВП на прирост населения

b. Прироста населения на динамику безработицы

c. Прирост людей с высшим образованием на рост промышленного производства

d. Прирост людей с высшим образованием на развитие науки

e. Прирост людей с высшим образованием на динамику доходов на душу населения

f. Динамику безработицы на динамику преступности

g. С помощью регрессионного анализа найдите зависимые переменные и поясните влияние на них независимых переменных.

h. С помощью функции predict() постройте прогноз по столбцу, соответствующему варианту.

Годы	Численность населения	Рост ВВП	Динамика безработицы	Динамика промышленного производства	Прирост людей, получивших очное высшее образование	Развитие науки (высокотехнологичных отраслей)	Динамика доходов на душу населения	Динамика преступности
	1	2	3	4	5	6	7	8
01.01.1990								
…								
01.01.2015								

Составитель (и): канд. техн. наук Вячкин Е.С.

(*фамилия, инициалы и должность преподавателя (ей)*)