

УТВЕРЖДАЮ

Первый проректор, проректор по
УР

_____ А.Е. Рудин

« 30 » 06 2020 года

Рабочая программа дисциплины

Б1.В.04 Интеллектуальный анализ данных

Учебный план: ФГОС 3++_2020-2021_09.04.03_ИИТА_ОО_ПИЭ_2-1-47.plx

Кафедра: **36** Информационных технологий

Направление подготовки:
(специальность) 09.04.03 Прикладная информатика

Профиль подготовки: Прикладная информатика в экономике
(специализация)

Уровень образования: магистратура

Форма обучения: очная

План учебного процесса

Семестр (курс для ЗАО)		Контактная работа обучающихся		Сам. работа	Контроль, час.	Трудоём- кость, ЗЕТ	Форма промежуточной аттестации
		Лекции	Практ. занятия				
2	УП	17	34	21	36	3	Экзамен
	РПД	17	34	21	36	3	
Итого	УП	17	34	21	36	3	
	РПД	17	34	21	36	3	

Рабочая программа дисциплины составлена в соответствии с федеральным государственным образовательным стандартом высшего образования по направлению подготовки 09.04.03 Прикладная информатика, утверждённым приказом Министерства образования и науки Российской Федерации от 19.09.2017 г. № 916

Составитель (и):

доктор технических наук, профессор

Пименов Виктор Игоревич

От кафедры составителя:

Заведующий кафедрой информационных технологий

Пименов Виктор Игоревич

От выпускающей кафедры:

Заведующий кафедрой

Пименов Виктор Игоревич

Методический отдел:

1 ВВЕДЕНИЕ К РАБОЧЕЙ ПРОГРАММЕ ДИСЦИПЛИНЫ

1.1 Цель дисциплины: Сформировать компетенции обучающегося в области применения методов, используемых для извлечения знаний из экономических данных.

1.2 Задачи дисциплины:

- дать представление об этапах извлечения знаний из баз данных;
- раскрыть индуктивные методы машинного обучения и статистические методы;
- проводить анализ применимости методов интеллектуального анализа данных при формировании знаний;
- выполнять оценку качества данных и сокращение многомерного пространства;
- выполнять классификацию многомерных объектов, прогнозировать показатели эффективности, осуществлять построение скоринговых карт с помощью методов Data Mining;
- использовать пакеты прикладных программ для анализа рыночной корзины.

1.3 Требования к предварительной подготовке обучающегося:

Предварительная подготовка предполагает создание основы для формирования компетенций, указанных в п. 2, при изучении дисциплин:

Когнитивные информационные технологии и системы

Математические методы и модели поддержки принятия решений

Основы научно-исследовательской деятельности

Эконометрика

2 КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ

ПКп-4 : Способен принимать эффективные проектные решения в условиях неопределенности и риска

Знать: информационный подход к моделированию; этапы обработки многомерных данных.

Уметь: анализировать данные, идентифицировать отклонения параметров и устанавливать влияние измеряемых параметров на показатели качества; выполнять глубокий анализ данных и извлекать знания.

Владеть: навыками многомерной классификации и дискриминантного анализа для извлечения знаний; построения диагностических правил и баз знаний.

3 РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ

Наименование и содержание разделов, тем и учебных занятий	Семестр (курс для ЗАО)	Контактная работа		СР (часы)	Инновац. формы занятий	Форма текущего контроля
		Лек. (часы)	Пр. (часы)			
Раздел 1. Подходы и методы приобретения знаний	2					О
Тема 1. Этапы извлечения знаний из баз данных (knowledge discovery in data). Направление Data Mining. Индуктивные методы машинного обучения и статистические методы. Технология экспертных систем. Ансамбли моделей. Практические занятия: Анализ предметной области и рекомендации по применению методов интеллектуального анализа данных.		4	6	4		
Тема 2. Статистические методы многомерного анализа данных. Согласование входных данных. Комплексный подход к использованию методов многомерного анализа данных. Факторный анализ. Регрессионный анализ. Практические занятия: Прогнозирование выходного показателя с помощью модели временного ряда.		4	6	4	ИЛ	
Раздел 2. Использование методов многомерного анализа данных при формировании знаний						
Тема 3. Методы сокращения многомерного пространства. Метод главных компонент. Канонический корреляционный анализ. Кластеризация переменных. Практические занятия: Сокращение многомерного пространства.		2	6	4		
Тема 4. Структуризация многомерных данных. Кластеризация объектов (обучение без учителя). Агломеративные и дивизимные методы иерархической кластеризации. Меры сходства / различия между объектами. Оценка расстояний между кластерами, принцип "средней связи", метод ближайшего соседа, метод наиболее удаленных соседей. Алгоритмы k-means и g-means. Дендрограмма. Выбор критерия качества кластеризации. Выбор числа кластеров. Практические занятия: Структуризация данных с использованием алгоритма g-means.		4	8	4		

Тема 5. Извлечение знаний из данных. Машинное обучение. Классификация объектов (обучение с учителем). Решающие правила, деревья решений, нейросетевые модели, эволюционные модели, области решений, семантические сети. Анализ рыночной корзины с помощью последовательных шаблонов. Геометрический подход, дискриминантный анализ. Методы локальной геометрии.		3	8	5	АС	
Практические занятия: Классификация многомерных объектов с помощью дискриминантных функций и дерева						
Итого в семестре (на курсе для ЗАО)		17	34	21		
Консультации и промежуточная аттестация (Экзамен)		2,5		33,5		
Всего контактная работа и СР по дисциплине		53,5		54,5		

4 КУРСОВОЕ ПРОЕКТИРОВАНИЕ

Курсовое проектирование учебным планом не предусмотрено

5. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

5.1 Описание показателей, критериев и системы оценивания результатов обучения

5.1.1 Показатели оценивания

Код компетенции	Показатели оценивания результатов обучения	Наименование оценочного средства
ПКп-4	Формулирует информационный поход к моделированию; перечисляет этапы обработки многомерных данных. Выполняет анализ данных, идентифицирует отклонения параметров и устанавливает влияние измеряемых параметров на показатели качества; выполняет глубинный анализ данных и извлекает знания с применением информационных технологий. Применяет инструментальные средства для извлечения знаний в ходе проведения многомерной классификации и дискриминантного анализа; построения диагностических правил и баз знаний.	Вопросы для устного собеседования. Практическое индивидуальное задание. Решение типовой задачи.

5.1.2 Система и критерии оценивания

Шкала оценивания	Критерии оценивания сформированности компетенций	
	Устное собеседование	Письменная работа
5 (отлично)	Полный, исчерпывающий ответ, явно демонстрирующий глубокое понимание предмета и широкую эрудицию в оцениваемой области. Критический, оригинальный подход к материалу.	Не предусмотрена
4 (хорошо)	Ответ полный, основанный на проработке всех обязательных источников информации. Подход к материалу ответственный. Ответ стандартный, в целом качественный, основан на всех обязательных источниках информации. Присутствуют небольшие пробелы в знаниях или несущественные ошибки.	Не предусмотрена
3 (удовлетворительно)	Ответ неполный, воспроизводит в основном только лекционные материалы, без самостоятельной работы с рекомендованной литературой. Демонстрирует понимание предмета в целом, без	Не предусмотрена

	углубления в детали. Присутствуют существенные ошибки или пробелы в знаниях по некоторым темам, незнание (путаница) важных терминов.	
2 (неудовлетворительно)	Неспособность ответить на вопрос без помощи экзаменатора. Незнание значительной части принципиально важных элементов дисциплины. Многочисленные грубые ошибки. Непонимание заданного вопроса. Неспособность сформулировать хотя бы отдельные концепции дисциплины. Попытка списывания, использования неразрешенных технических устройств или пользования подсказкой другого человека (вне зависимости от успешности такой попытки).	Не предусмотрена

5.2 Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности

5.2.1 Перечень контрольных вопросов

№ п/п	Формулировки вопросов
Семестр 2	
1	Анализ рыночной корзины с помощью последовательных шаблонов.
2	Модель логистической регрессии. Построение скоринговой карты.
3	Оценка качества данных. Восстановление пропущенных значений, редактирование аномальных значений и спектральная обработка в целях сглаживания данных.
4	Инструментальные средства, используемые для интеллектуального анализа данных.
5	Методы локальной геометрии.
6	Геометрический подход, дискриминантный анализ.
7	Семантические сети.
8	Эволюционные модели.
9	Нейросетевые модели.
10	Модель дерева решений.
11	Классификация объектов (обучение с учителем). Решающие правила.
12	Извлечение знаний из данных. Машинное обучение.
13	Дендрограмма. Выбор критерия качества кластеризации. Выбор числа кластеров.
14	Алгоритмы k-means и g-means.
15	Метод ближайшего соседа, метод наиболее удаленных соседей.
16	Меры сходства / различия между объектами. Оценка расстояний между кластерами, принцип "средней связи".
17	Агломеративные и дивизимные методы иерархической кластеризации.
18	Структуризация многомерных данных. Кластеризация объектов (обучение без учителя).
19	Кластеризация переменных.
20	Канонический корреляционный анализ.
21	Методы сокращения многомерного пространства. Метод главных компонент.
22	Регрессионный анализ.
23	Факторный анализ.
24	Комплексный подход к использованию методов многомерного анализа данных.
25	Согласование входных данных.
26	Статистические методы многомерного анализа данных.
27	Ансамбли моделей.
28	Технология экспертных систем.
29	Направление Data Mining. Индуктивные методы машинного обучения и статистические методы.
30	Этапы извлечения знаний из баз данных (knowledge discovery in data).

5.2.2 Типовые тестовые задания

Не предусмотрено

5.2.3 Типовые практико-ориентированные задания (задачи, кейсы)

Типовые практико-ориентированные задания находятся в Приложении к данной РПД.

5.3 Методические материалы, определяющие процедуры оценивания знаний, умений, владений (навыков и (или) практического опыта деятельности)

5.3.1 Условия допуска обучающегося к промежуточной аттестации и порядок ликвидации академической задолженности

Проведение промежуточной аттестации регламентировано локальным нормативным актом СПбГУПТД «Положение о проведении текущего контроля успеваемости и промежуточной аттестации обучающихся»

5.3.2 Форма проведения промежуточной аттестации по дисциплине

Устная Письменная Компьютерное тестирование Иная

5.3.3 Особенности проведения промежуточной аттестации по дисциплине

В течение семестра выполняются контрольные работы.

Время на подготовку ответа экзаменационного билета составляет 30 минут.

Время на выполнение практического задания экзаменационного билета с применением вычислительной техники составляет 20 минут.

При проведении экзамена не разрешается пользоваться учебными материалами.

Экзамен проводится в компьютерном классе.

6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1 Учебная литература

Автор	Заглавие	Издательство	Год издания	Ссылка
6.1.1 Основная учебная литература				
Пименов В. И., Пименов И. В.	Информационный менеджмент	СПб.: СПбГУПТД	2019	http://publish.sutd.ru/tp_ext_inf_publish.php?id=201911
Пименов В.И., Суздамов Е.Г., Воронов М.В.	Методы обработки информации в научных исследованиях	СПб.: СПбГУПТД	2015	http://publish.sutd.ru/tp_ext_inf_publish.php?id=2641
Пальмов С. В.	Интеллектуальный анализ данных	Самара: Поволжский государственный университет телекоммуникаций и информатики	2017	http://www.iprbookshop.ru/75376.html
6.1.2 Дополнительная учебная литература				
Пименов В. И., Пименов И. В.	Интеллектуальный анализ данных	СПб.: СПбГУПТД	2017	http://publish.sutd.ru/tp_ext_inf_publish.php?id=201748
Нестеров С. А.	Интеллектуальный анализ данных средствами MS SQL Server 2008	Москва: Интернет-Университет Информационных Технологий (ИНТУИТ)	2016	http://www.iprbookshop.ru/62813.html
Уэс Маккинли, Слинкин А. А.	Python и анализ данных	Саратов: Профобразование	2017	http://www.iprbookshop.ru/64058.html

6.2 Перечень профессиональных баз данных и информационно-справочных систем

Федеральная служба государственной статистики [Электронный ресурс]. – Режим доступа: <http://www.gks.ru>.

Информационная система «Единое окно доступа к образовательным ресурсам» [Электронный ресурс]. URL: <http://window.edu.ru>.

Электронно-библиотечная система IPRbooks [Электронный ресурс]. URL: <http://www.iprbookshop.ru>.

6.3 Перечень лицензионного и свободно распространяемого программного обеспечения

Mathcad Education – University Edition Term

MicrosoftOfficeProfessional

Microsoft Windows

Microsoft Windows Home Russian Open No Level Academic Legalization Get Genuine (GGK) + Microsoft Windows Professional (Pro – профессиональная) Russian Upgrade Open No Level Academic

6.4 Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

Аудитория	Оснащение
Лекционная аудитория	Мультимедийное оборудование, специализированная мебель, доска
Компьютерный класс	Мультимедийное оборудование, компьютерная техника с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду

Приложение

рабочей программы дисциплины “Интеллектуальный анализ данных”

по направлению подготовки 09.04.03 Прикладная информатика

наименование ОП (профиля): Прикладная информатика в экономике

5.2.3 Типовые практико-ориентированные задания

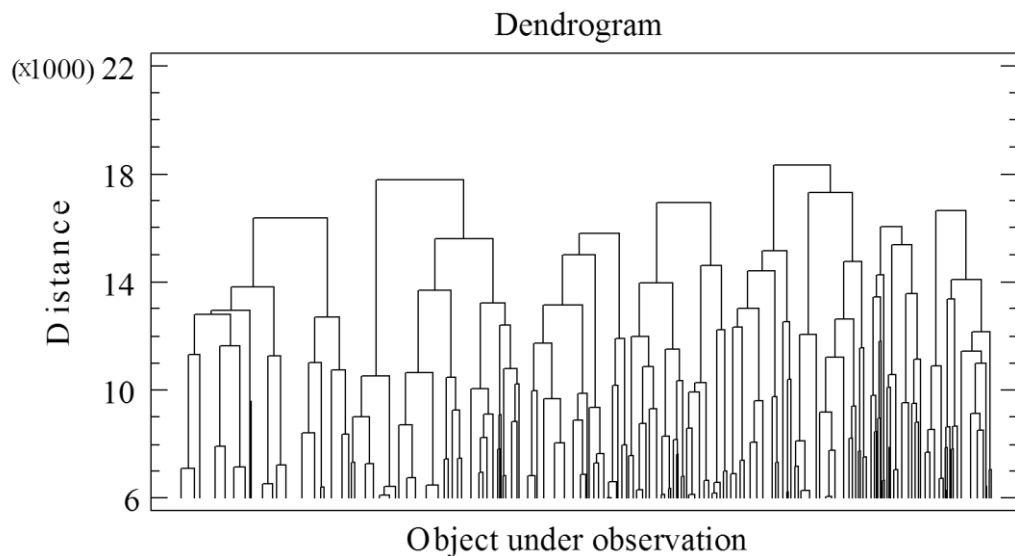
№ п/п	Условия типовых практико-ориентированных заданий
Семестр 2	
1	<p>Предположим, все множество знаний предметной области структурируется на основе графа, представленного на рисунке. Какую модель представления знаний целесообразно выбрать для формального описания знаний? Почему</p> <div style="text-align: center;"> </div>
2	<p>Найдена регрессионная модель:</p> $y = \text{logit}(p) = 1,5 - 0,6 \cdot x_1 + 0,4 \cdot x_2 - 0,3 \cdot x_3.$ <p>Имеется набор новых наблюдений $(x_1, x_2, x_3) = (1, 0, 1)$. Используя логистическую модель, оцените вероятность появления выходного значения 1, то есть $P(y = 1)$.</p> <p><i>Решение</i></p> <p>Вычислим соответствующее логит-преобразование: $\text{logit}(p) = 1,5 - 0,6 \cdot 1 + 0,4 \cdot 0 - 0,3 \cdot 1 = 0,6$. То есть $\text{log}_e \left(\frac{p_j}{1 - p_j} \right) = 0,6$. Вероятность появления значения 1: $P(Y = 1) = e^{0,6} / (1 + e^{0,6}) = 0,64$.</p> <p>Таким образом, можно заключить, что выходное значение $Y = 1$ более вероятно, чем $Y = 0$.</p>
3	<p>На основе кредитных историй заемщиков в среде Deducor построены профили кластеров, фрагмент которых приведен на рисунке.</p>

		Клusters	
		0	Итого
		28 (18,8%)	
+ Поля	Показатели		
9.0 Срок проживания в данной местности, лет	Значимость	100,0%	100,0%
	Доверительный интервал		
	Среднее	26,21428571	17,85234899
	Стандартн. откл.	7,40548917	7,775025714
	Стандартн. ошиб.	1,399505906	0,6369549248
9.0 Количество лет	Значимость	100,0%	100,0%
9.0 Срок работы на	Значимость	100,0%	100,0%
9.0 Среднемесячный доход, руб	Значимость	94,3%	92,5%
	Доверительный интервал		
	Среднее	7910,714286	9516,778523
	Стандартн. откл.	2600,048331	4288,740801
	Стандартн. ошиб.	491,3629486	351,3473363
9.0 Срок ссуды, мес	Значимость	54,5%	42,4%
9.0 Размер ссуды, руб	Значимость	44,5%	29,7%

Дайте интерпретацию кластеру с номером 0.

4

В результате многомерного анализа данных в пакете Statgraphics получена дендрограмма кластеров. Сделать вывод о количестве выделенных кластеров.



5

На рисунке представлена матрица расстояний самоорганизующихся карт Кохонена. Дайте интерпретацию матрицы.

