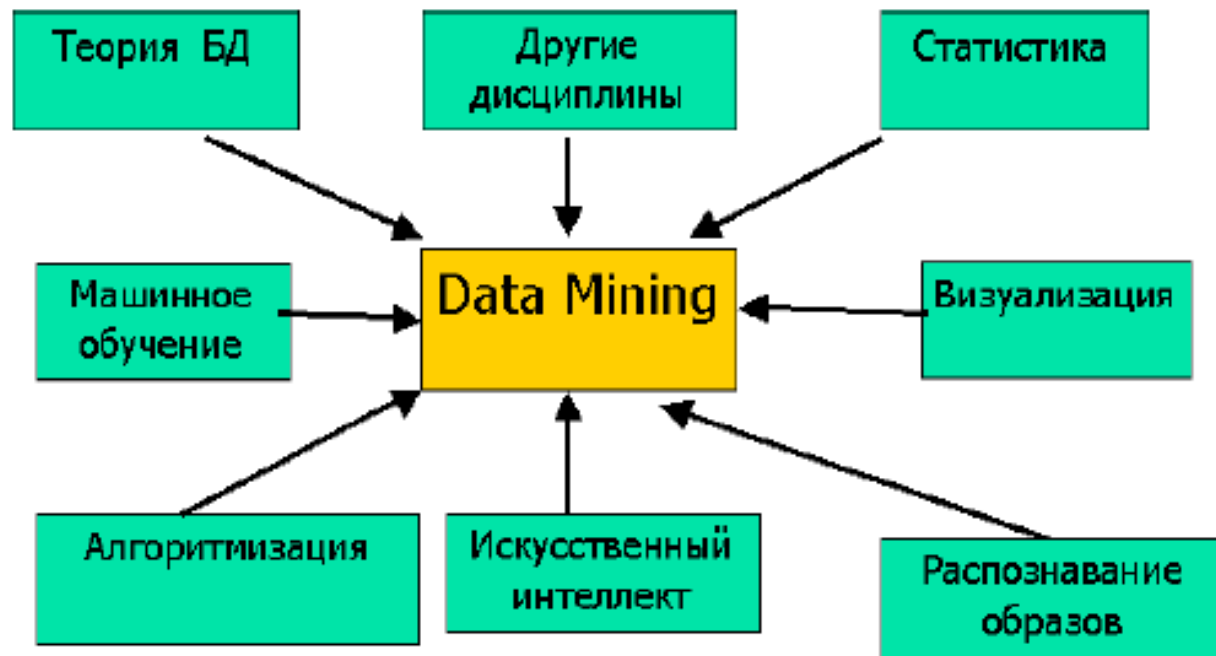


Интеллектуальный анализ данных



1. Сравнение статистики, машинного обучения и Data Mining

- Статистика
 - Более, чем Data Mining, базируется на теории.
 - Более сосредотачивается на проверке гипотез.
- Машинное обучение
 - Более эвристично.
 - Концентрируется на улучшении работы агентов обучения.
- Data Mining.
 - Интеграция теории и эвристик.
 - Сконцентрирована на едином процессе анализа данных, включает очистку данных, обучение, интеграцию и визуализацию результатов.

Шифр навчальної дисципліни	Назва навчальної дисципліни	Назва блока змістового модуля	Шифр блока змістових модулів	Назва теми (змістового модуля)	Шифр теми (змістового модуля)
3.04	Організація баз даних та знань	Моделювання даних	3.04.01	Системи баз даних. Основні поняття й архітектура	3.04.01.01
				Моделі даних	3.04.01.02
				Реляційна модель даних	3.04.01.03
				Теорія нормалізації реляційної моделі даних	3.04.01.04
		Мови запитів	3.04.02	Мова SQL	3.04.02.01
				Мова QBE	3.04.02.02
		Проектування та захист баз даних	3.04.03	Проектування баз даних	3.04.03.01
				Цілісність даних	3.04.03.02
				Захист баз даних	3.04.03.03
				Навігаційна обробка даних	3.04.03.04
		Класифікація баз даних	3.04.04	Розподілені бази даних	3.04.04.01
				Паралельні бази даних	3.04.04.02
				Дедуктивні бази даних	3.04.04.03
				Об'єктно-орієнтовані бази даних	3.04.04.04
				Бази даних в Інтернеті	3.04.04.05
				Бази знань	3.04.04.06
3.05	Інтелектуальний аналіз даних	Основи інтелектуального аналізу даних	3.05.01	Методи первісної обробки даних	3.05.01.01
				Методи дослідження структури даних: візуалізація та автоматичне групування даних	3.05.01.02
		Методи використання навчальної інформації	3.05.02	Кореляційний і регресійний аналіз даних. Множинний регресійний аналіз.	3.05.02.01
				Лінійна множинна регресійна модель. Перевірка адекватності моделі	3.05.02.02
				Нелінійне оцінювання параметрів	3.05.02.03
		Методи багатомірного розвідувального аналізу	3.05.03	Кластерний аналіз. Ієрархічна та секційна кластеризації	3.05.03.01
				Методи кластеризації: процедура Мак-Кіна, метод k-методів, сітчасті методи	3.05.03.02
				Растрова кластеризація об'єктів	3.05.03.03
				Лінійний дискримінантний аналіз. Побудова канонічних та класифікаційних функцій	3.05.03.04
		Методи класифікації та прогнозування	3.05.04	Дерева рішень	3.05.04.01
				Методи опорних векторів, «найближчого сусіда», Байеса	3.05.04.02
				Аналіз багатомірних угруповань	3.05.04.03
				Статистична обробка тимчасових рядів і прогнозування	3.05.04.04
				Класифікація об'єктів у випадку невідомих розподілень даних	3.05.04.05
				Методи оцінювання помилок класифікації	3.05.04.06

Шифр навчальної дисципліни	Назва навчальної дисципліни	Назва блока змістового модуля	Шифр блока змістових модулів	Назва теми (змістового модуля)	Шифр теми (змістового модуля)
3.06	Веб-технології та веб-дизайн	Методи пошуку шаблонів даних	3.05.05	Асоціаційні правила. Послідовне відображення шаблонів даних	3.05.05.01
				Метод Apriori, побудова FP-дерев пошуку шаблонів даних.	3.05.05.02
				Min-max асоціації у базах даних	3.05.05.03
				Побудова hash-дерев	3.05.05.04
				Розробка OLAP-кубів під час аналізу багатомірних даних у великих БД.	3.05.05.05
				Способи та методи візуального відображення даних	3.05.05.06
		OLAP і Data Mining	3.05.06	Методи, стадії, задачі Data Mining	3.05.06.07
				Упровадження Data Mining, OLAP і сховищ даних у СППР	3.05.06.08
				Процес Data Mining	3.05.06.01
				Стандарти Data Mining	3.05.06.02
				Інструменти Data Mining	3.05.06.03
				Структура і принципи Веб	3.06.01.01
		Основи Веб	3.06.01	Уведення в клієнт-серверні технології Веб	3.06.01.02
				Протокол HTTP	3.06.01.03
				Клієнтські сценарії та застосування	3.06.01.04
				Серверні веб-застосування	3.06.01.05
				JavaScript. Програмна взаємодія з HTML документами на основі DOM API	3.06.02.01
		Веб-програмування	3.06.02	Мови розроблення сценаріїв Perl, PHP, JSP	3.06.02.02
				Розробка CGI-застосувань на Perl, PHP, JSP	3.06.02.03
				Основи розробки веб-застосувань з допомогою ASP.NET, J2EE	3.06.02.04
				Інтерфейси взаємодії веб-застосувань з СКБД	3.06.02.05
				Веб-сервіси та мови їх описування	3.06.02.06
				Мови описування схем XML	3.06.03.01
		Основи XML	3.06.03	DOM XML. Перетворення XML-документів	3.06.03.02
				Програмна обробка XML-документів з допомогою XML DOM	3.06.03.03
				Форматування і перетворення XML-документа з допомогою CSS та XSL. XSLT перетворення XML-документа	3.06.03.04
				Інтеграція та взаємодія у веб-мережі	3.06.04.01
		Веб-портالي	3.06.04	Розробка веб-служби в ASP.NET, J2EE	3.06.04.02
				Розробка веб-контента. CMS/CMF	3.06.04.03
				Розробка RSS-джерел і RSS-рідерів	3.06.04.04
				Уведення в технологію AJAX. Розробка мобільних веб-застосувань	3.06.05.01
		Технологія AJAX	3.06.05	Реалізація асинхронної взаємодії браузера з веб-сервером з допомогою технології AJAX	3.06.05.02

Змістовий модуль 1. Основи інтелектуального аналізу даних.

Тема 1. Методи первісної обробки даних.

Тема 2. Методи дослідження структури даних: візуалізація та автоматичне групування даних.

Змістовий модуль 2. Методи використання навчальної інформації.

Тема 1. Кореляційний і регресійний аналіз даних. Множинний регресійний аналіз.

Тема 2. Лінійна множинна регресійна модель. Перевірка адекватності моделі.

Тема 3. Нелінійне оцінювання параметрів.

Змістовий модуль 3. Методи багатомірного розвідувального аналізу.

Тема 1. Кластерний аналіз. Ієрархічна та секційна кластеризації.

Тема 2. Методи кластеризації: процедура Мак-Кина, метод k-means, сітчасті методи.

Тема 3. Растрова кластеризація об'єктів.

Тема 4. Лінійний дискримінантний аналіз. Побудова канонічних та класифікаційних функцій.

Змістовий модуль 4. Методи класифікації та прогнозування.

Тема 1. Дерева рішень.

Тема 2. Методи опорних векторів, «найближчого сусіда», Байеса.

Тема 3. Аналіз багатомірних угруповань.

Тема 4. Статистична обробка тимчасових рядів і прогнозування.

Тема 5. Класифікація об'єктів у випадку невідомих розподілень даних.

Тема 6. Методи оцінювання помилок класифікації.

1. Програма дисципліни

Змістовий модуль 5. Методи пошуку шаблонів даних.

Тема 1. Асоціаційні правила. Послідовне відображення шаблонів даних.

Тема 2. Метод Apriori, побудова FP-дерев пошуку шаблонів даних.

Тема 3. Min-max асоціації у базах даних.

Тема 4. Побудова hash-дерев.

Тема 5. Розробка OLAP-кубів під час аналізу багатомірних даних у великих БД.

Тема 6. Способи та методи візуального відображення даних.

Змістовий модуль 6. OLAP і Data Mining.

Тема 1. Методи, стадії, задачі Data Mining.

Тема 2. Упровадження Data Mining, OLAP і сховищ даних у СППР.

Тема 3. Процес Data Mining.

Тема 4. Стандарти Data Mining.

Тема 5. Інструменти Data Mining.

1. Методы первичной обработки данных

- 1. Проблемы количественных измерений**
- 2. Этапы и процедуры первичной обработки данных**
- 3. Методы математической и статистической обработки данных**
- 4. Проверка гипотез**
- 5. Табличное и графическое представление данных**



**Способы
статистического
наблюдения**

Непосредственное
наблюдение

Документальное
наблюдение или
документальная
запись

Опрос

1. Методы первичной обработки данных

1. История развития статистики как науки.
2. Основные понятия и методы статистики.
3. Структура отраслей статистической науки.*
4. Организация государственной системы статистики РФ.*



1. Методы первичной обработки данных

	Атрибуты				
	Код клиента	Возраст	Семейное положение	Доход	Класс
Объекты	1	18	Single	125	1
	2	22	Married	100	1
	3	30	Single	70	1
	4	32	Married	120	1
	5	24	Divorced	95	2
	6	25	Married	60	1
	7	32	Divorced	220	1
	8	19	Single	85	2
	9	22	Married	75	1
	10	40	Single	90	2

1. Методы первичной обработки данных

Измерение - процесс присвоения чисел характеристикам изучаемых объектов согласно определенному правилу.

В процессе подготовки данных измеряется не сам объект, а его характеристики.

Шкала - правило, в соответствии с которым объектам присваиваются числа.

1. Методы первичной обработки данных

Измерение - процесс присвоения чисел характеристикам изучаемых объектов согласно определенному правилу.

В процессе подготовки данных измеряется не сам объект, а его характеристики.

Шкала - правило, в соответствии с которым объектам присваиваются числа.

1. Методы первичной обработки данных

	Атрибуты				
	Код клиента	Возраст	Семейное положение	Доход	Класс
Объекты	1	18	Single	125	1
	2	22	Married	100	1
	3	30	Single	70	1
	4	32	Married	120	1
	5	24	Divorced	95	2
	6	25	Married	60	1
	7	32	Divorced	220	1
	8	19	Single	85	2
	9	22	Married	75	1
	10	40	Single	90	2

1. Методы первичной обработки данных

Измерение - процесс присвоения чисел характеристикам изучаемых объектов согласно определенному правилу.

В процессе подготовки данных измеряется не сам объект, а его характеристики.

Шкала - правило, в соответствии с которым объектам присваиваются числа.

1. Методы первичной обработки данных

	Атрибуты				
	Код клиента	Возраст	Семейное положение	Доход	Класс
Объекты	1	18	Single	125	1
	2	22	Married	100	1
	3	30	Single	70	1
	4	32	Married	120	1
	5	24	Divorced	95	2
	6	25	Married	60	1
	7	32	Divorced	220	1
	8	19	Single	85	2
	9	22	Married	75	1
	10	40	Single	90	2

1. Методы первичной обработки данных

Порядковая шкала (ordinal scale) - шкала, в которой числа присваивают объектам для обозначения относительной позиции объектов, но не величины различий между ними.

Интервальная шкала (interval scale) - шкала, разности между значениями которой могут быть вычислены, однако их отношения не имеют смысла.

Интервальная шкала (interval scale) - шкала, разности между значениями которой могут быть вычислены, однако их отношения не имеют смысла.

Дихотомическая шкала (dichotomous scale) - шкала, содержащая только две категории.

Методы исследования структуры данных: визуализация и автоматическая группировка данных

Data Mining - это процесс поддержки принятия решений, основанный на поиске в данных скрытых закономерностей (шаблонов информации) [3].

Технологию Data Mining достаточно точно определяет Григорий Пиатецкий-Шапиро (Gregory Piatetsky-Shapiro) - один из основателей этого направления:

Data Mining - это процесс обнаружения в сырых данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности.

Суть и цель технологии Data Mining можно охарактеризовать так: это технология, которая предназначена для поиска в больших объемах данных неочевидных, объективных и полезных на практике закономерностей.

1. Методы первичной обработки данных

Неочевидных - это значит, что найденные закономерности не обнаруживаются стандартными методами обработки информации или экспертным путем.

Объективных - это значит, что обнаруженные закономерности будут полностью соответствовать действительности, в отличие от экспертного мнения, которое всегда является субъективным.

Практически полезных - это значит, что выводы имеют конкретное значение, которому можно найти практическое применение.