

Конспект за курса
“Извличане на закономерности от данни”
2011 г.

1. **Какво е ИЗД?** Модел на процеси в ИЗД. Типове задачи в ИЗД.
2. **Разбиране на данни.** Основни статистически характеристики на данни и тяхна визуализация. Изследване на зависимости между атрибути.
3. **Подготовка на данни.** “Изчистване” на данни. Интеграция на данни. Трансформация на данни (нормализация и дискретизация).
4. **Намаляване на обема на данни.** Намаляване на размерността на данни. Компресия на данни. Моделиране и подбор на данни. Обобщение чрез понятийни йерархии.
5. **Основи на клъстерния анализ.** Типове данни, използвани при клъстерния анализ. Категоризация на основни методи за клъстеризация.
6. **Основни методи за нейерархична клъстеризация.** Клъстеризация чрез разделяне (K-Means, Разполовяващ k-Means, K-Modes, K-Prototypes, k-Medoids). Клъстеризация чрез разделяне в големи бази данни
7. **Основни методи за йерархична клъстеризация.** Агломеративна и разделяща клъстеризация. Алгоритмите BIRCH и CURE
8. **Методи за клъстеризация, базирани на плътността.** Алгоритмите DBSCAN, DENCLUE и клъстеризация, азирана на мрежи.
9. **Методи за клъстеризация, използващи сходство, базирано на най-близки съседи.** SNN – сходство, SNN-плътност и тяхното използване за клъстеризация.
10. **Оценяване на клъстери.** Класификация на подходи за оценяване на клъстери. Направляване и ненаправлявани подходи. Оценяване на йерархична клъстеризация.
11. **Задача за регресия.** Съществуващи регресионни методи – статистически и от машинно самообучение.
12. **Регресионни дървета.** Алгоритъм за ефективно създаване на LS-регресионни дървета.
13. **Избягване на преспецификацията при регресионни дървета и локални регресионни дървета.** Едно-стъпкови методи за завършващото подрязване на регресионни дървета. Дву-стъпкови методи за завършващото подрязване на регресионни дървета. Локални регресионни дървета.
14. **Извличане на асоциативни правила.** Основни понятия. Класификация на асоциативните правила. Извличане на едномерни булеви асоциативни правила от транзакционни бази данни – алгоритъм Apriori.
15. **Подобряване на ефективността на Apriori.** Намиране на често срещани артикулни множества без генериране на кандидати.
16. **Извличане на многонивови асоциативни правила.** Извличане на многонивови асоциативни правила от транзакционни бази данни.
17. **Извличане на многомерни асоциативни правила от релационни бази данни.** Мерки за интересност.
18. **Откриване на аномалии.** Класификация на подходи. Статистически подходи. Класификационни подходи. Подходи, базирани на разстояние и плътност. Подходи, базирани на клъстеризация.

19. **Организация на данни, складове данни и OLAP технология.** Йерархията на паметта. Индексни структури. Многомерно индексирани. Модел на многомерни данни. Три-нивова архитектура на складове данни. Имплементация на складове данни.
20. **Анализ на текста.** Индексиране. Лексическа обработка. Подреждане, базирано на съдържание. Анализ на латентна семантика. Намиране на асоциации между думи.