

СОФИЙСКИ УНИВЕРСИТЕТ "СВ. КЛИМЕНТ ОХРИДСКИ"  
ФАКУЛТЕТ ПО МАТЕМАТИКА И ИНФОРМАТИКА

учебна година: 2009/2010

семестър: летен  
(зимен, летен)

|  |
|--|
| <b>наименование на дисциплината:</b><br>"Извличането на знания от данни" |
| <b>хорариум: 45+30 вид на дисциплината:</b><br><b>избираема</b>          |
| <b>специалност:</b><br><b>курс: 4-ти</b>                                 |
| <b>лектор: доцент д-р Иван Койчев</b>                                    |

1. Кратка анотация на дисциплината

Извличането на знания от данни (Data Mining още известни и като Knowledge Discovering in Data) една много активно развиващите се области с множество реални приложения. Курсът си поставя за цел да запознае студентите с основните методи и алгоритми на Извличането на знания от данни както и примери за тяхното приложение.

По време на лабораторните занятия студентите ще се запознаят със съответните отворени среди и ще получат задания за самостоятелна работа (проект), който ще представят пред колегите си в края на курса.

2. Предварителни изисквания към студентите (отнася се само за избираемите дисциплини)
3. Форма на проверка на знанията и уменията и начин на формиране на оценката по дисциплината

|                             | % от оценката |
|-----------------------------|---------------|
| Текуща оценка               | 65%           |
| – курсова работа            | 55%           |
| – котролна работа           | 0%            |
| – активно учатие в часовете | 5%            |
| – присъствие в час          | 5%            |
| Изпит                       | 35%           |
| – практически (задачи)      | 0             |
| – теоретически              | 35%           |

#### 4. Тематичен план (конспект) на дисциплината

### КОНСПЕКТ

| №   | Тема   |
|-----|--|
| 1.  | Въведение. Обучение на машини и извличане на знания от данни.  |
| 2.  | Данни. Понятия. Примери. Атрибути. Подготовка на данни за анализ.  |
| 3.  | Представяне на извлечените знания.   |
| 4.  | Оценка на качеството на извлечените знания   |
| 5.  | Класификация: Учене на класификатори. Дърво на Решенията.  |
| 6.  | Класификация: Статистически методи. Наивен Бейсов Модел.   |
| 7.  | Класификация: Групови решения. Итеративно подобрчване на моделите.   |
| 8.  | Класификация: Учене основано на примери. Локални и глобални модели.  |
| 9.  | Методи за избор на значими атрибути.   |
| 10. | Класификация: Машини с поддържащи вектори (SVM)  |
| 11. | Откриване на асоциации в данни: основни понятия и алгоритми.   |
| 12. | Откриване на асоциации в данни: други методи. Основни приложения.  |
| 13. | Клъстеризация: основни понятия и алгоритми: К-средни и негови производни. Измерване на качеството на клъстерите. |
| 14. | Клъстеризация: Агломеративни алгоритми.  |
| 15. | Изпозване на Невронни Мрежи. Персептрон. Winnow.   |
| 16. | Откриване на аномалии.   |
| 17. | Визуализация на извлечените знания. Прокетиране на многомерни пространства в двумерни.                           |
| 18. | Откриване на знания в текст и мултимедия.  |
| 19. | Откриване на знания в уеб  |

#### 5. Литература

1. Berkhin P, Survey of Clustering Data Mining Techniques
2. Hipp J., Ulrich Guntzer, Gholamreza Nakhaeizadeh: Algorithms for Association Rule Mining - A General Survey and Comparison. SIGKDD Explorations 2(1): 58-64 (2000)
3. Murthy S., Automatic Construction of Decision Trees from Data: A Multi-Disciplinary Survey. Journal of Data Mining and Knowledge Discovery, vol. 2, num. 4, 1998
4. Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall, Second Edition, 2003.
5. Tom Mitchell, Machine Learning, McGraw Hill, 1997.
6. Ruth Dilly, Data Mining - An Introduction. The Queen's University of Belfast - OHP Slide Material, Student Notes.
7. Kurt Thearling - "An Introduction to Data Mining" -
8. Andrew Moore. Statistical Data Mining Tutorials
9. Данаил Дочев, Здравко Марков, Геннадий Агре (2002) Увод в машинното самообучение, серия "Лекции по основи на компютърната наука", Софтех, София.