

D – Charakteristika studijního předmětu					
Název studijního předmětu	Data mining				
Typ předmětu	povinně volitelný		doporučený ročník / semestr	1/L	
Rozsah studijního předmětu	2 + 2	hod. za týden	4	kreditů	5
Jiný způsob vyjádření rozsahu	—				
Způsob zakončení	zkouška		Forma výuky	př + cv	
Další požadavky na studenta					
Absolvování předepsaných úloh, odevzdání a obhájení individuální semestrální práce, znalost odpřednášené látky a uspokojivé výsledky průběžné kontroly. Zkouška je písemná a ústní.					
Vyučující					
RNDr. Klára Císařová, Ph.D.					
Stručná anotace předmětu					
Předmět se soustřeďuje na typické úlohy a postupy dataminingu (DM). Učí základy DM technik, jako jsou klasifikační, segmentační, asociační a další algoritmy, které pomáhají v praxi k využití rozsáhlých dat. Těmito postupy se studenti učí predikovat významné veličiny za účelem optimalizace rozhodování v praxi, hledat anomálie, hledat vztahy a vzory či jiné skryté informace. DM úlohy budou řešeny metodologií CRISP-DM ve vhodném softwarovém nástroji. Studenti se naučí všechny kroky DM od formulace problému, přes získání dat, přípravy a zpracování dat pro modelování, návrh a ověření modelu, až po nasazení řešení do praxe. Důraz bude kladen na porozumění a interpretaci výsledků.					
Témata přednášek:					
<div>1. Proces dobývání znalostí – historie, vymezení cílů, přehled metodologií.</div> <div>2. Dělení dataminingových úloh, typické úlohy – CRM péče o zákazníka, získávání nových zákazníků, predikce odchodů zákazníků ke konkurenci (churn), marketingová úspěšnost obchodních kampaní apod.</div> <div>3. Odhalování podvodů – fraud úlohy, kreditní riziko, behaviorální skórování pro hodnocení rizikovitosti.</div> <div>4. Příprava dat, porozumění datům, popis datové množiny, příprava datové matice, výběr dat a jejich čištění, konstrukce a slučování datových zdrojů, typová homogenita, formátování dat.</div> <div>5. Klasifikační algoritmy jako nástroje predikce vycházející z historických dat. Rozhodovací stromy, algoritmy C&RT, C5.0, CHAID&QUEST. Převod stromu na pravidla, prořezávání stromů.</div> <div>6. Diskriminační analýza – klasifikace případů do tříd, skórování.</div> <div>7. Segmentační algoritmy – odhalení neobvyklých struktur v datech nasazením algoritmů seskupování K-Means, Two Step, Anomaly.</div> <div>8. Asociační algoritmy – hledání asociačních pravidel, model Apriori, Carma, statistiky implikací, predikční model.</div> <div>9. Základy neuronových sítí pro zpracování kategorizovaných i číselných proměnných, použití v případech, kdy klasické lineární metody neposkytují očekávané výsledky.</div> <div>10. Analýza a predikce časových řad pomocí modelů DM, příprava dat, doplnění chybějících hodnot, difference, sezónní difference, klouzavé průměry a mediány, vyhlazování časových řad.</div> <div>11. Modelování a evaluace řešení, zavádění DM řešení do praxe, zařazení skórovacích procesů do rozhodovacího firemního workflow.</div> <div>12. Webmining, Textmining.</div> <div>13. Rozbor typických úloh data miningu.</div> <div>14. Rozbor typických úloh data miningu.</div>					
Informace ke kombinované nebo distanční formě					
Rozsah konzultací (soustředění)		hodin za týden			
Rozsah a obsahové zaměření individuálních prací studentů a způsob kontroly					
Studijní literatura a studijní pomůcky					
YIN, Yong, Ikou KAKU, Jiafu TANG a JianMing ZHU. <i>Data Mining</i> . Springer. 2011. ISBN 978-1849963374.					
BERKA, Petr. <i>Dobývání znalostí z databází</i> . Praha: Academia, 2003, ISBN 80-200-1062-9.					
HENDL, Jan. <i>Přehled statistických metod zpracování dat: analýza a metaanalýza dat</i> . Vyd. 2., opr. Praha: Portál, 2006, 583 s. ISBN 80-7367-123-9.					
PARR RUD, Olivia. <i>Data Mining</i> . Computer Press. 2002. ISBN 8072265776.					
Kurz na portálu ALS s podpůrnými programy pro experimentování a hlubšímu porozumění vybraným algoritmům, samostatné práce s podporou KNIME, WEKA, IBM SPSS MODELERU.					