

## UPLATNENIE TECHNOLOGIE DATA MINING NAD DATABÁZOU AIS2

### APPLICATION OF DATA MINING TECHNOLOGY OVER THE DATABASE AIS2

Jolana Gubalová

Univerzita Mateja Bela v Banskej Bystrici

Ekonomická fakulta, Katedra kvantitatívnych metód a informačných systémov,  
jolana.gubalova@umb.sk

#### Abstrakt

*Uplatnenie analytických nástrojov sa v priebehu posledných rokov stalo kľúčovým trendom v procese prijímania včasných a správnych rozhodnutí. Nielen svetové spoločnosti, ale aj slovenské firmy postupne implementujú do svojich informačných systémov technológie umožňujúce rýchlo a efektívne analyzovať podnikové údaje. Náš príspevok sa zaoberá aplikáciou technológie Data Mining, integrovanej v systéme MS SQL Server v riadení verejnej vysokej školy. Objektom skúmania je súbor reálnych údajov extrahovaných z databázy Akademického informačného systému (AIS2) Univerzity Mateja Bela. Príspevok opisuje proces vytvárania modelov na analýzu výsledkov hodnotení vybraných predmetov a zisťovanie asociácií v rámci výberových predmetov. Stanovuje možnosti sprístupnenia výsledkov analýz bežnému používateľovi v prehľadnej forme, pri použití štandardných kancelárskych aplikácií. Príspevok sumarizuje a vyhodnocuje poznatky získané pri procese zostavovania analytických modelov a formuluje odporúčania pre efektívnejšiu prácu s AIS2.*

#### Abstract

*The application of analytical tools in recent years become a key trend in the adoption of timely and correct decisions. Not only global companies, as well as Slovak company, gradually implemented in its information technology systems to respond rapidly and effectively analyze business data. Our paper deals with the application of data mining technology, integrated in MS SQL Server in the management of public higher education. The object of the study is a set of real data extracted from the database of the Academic Information System (AIS2), Matej Bel University. The paper describes the process of creating a model for analyzing the results of evaluations of selected objects and to detect associations within the framework of the course. It establishes the possibility of opening the results of analysis of normal user in clear form, using standard office applications. The article summarizes and evaluates the knowledge acquired in the compilation of analytical models and formulates recommendations for more effective work with AIS2.*

#### Kľúčové slová

*Databázový relačný systém, akademický informačný systém, analytické nástroje, Data Mining, Business Intelligence*

## 1. Úvod

Veľké objemy dát z podnikového prostredia, finančných transakcií, reklamných analýz, z vedeckého skúmania a množstvo iných druhov dát, predstavujú informácie v menej alebo viac štruktúrovanej podobe. Ich nárast je navyše poháňaný technologickým rozvojom a spoločenským trendom integrácie informačných technológií do bežného života v čoraz väčšej miere. Na zvládnutie takéhoto množstva dát je potrebný systematický prístup podporený špecifickými technológiami. Práve Data Mining je jednou z takýchto technológií. Už mnohokrát spomínanou frázou v súvislosti s Data Miningom je: “Dnes už nie je problémom

nedostatok dát, v skutočnosti sme zahltení dátami z viacerých zdrojov. Ozajstným problémom je nedostatok analytikov schopných nájsť vedomosti v dátach.” (LAROSE, 2005, s.4).

Presná definícia data miningu neexistuje. Môžeme len interpretovať niektorých autorov. Krátka definícia od Hana (2006, s.5) znie „Data Mining je extrakciou alebo dolovaním znalostí z veľkého objemu dát“. Larose (2005, s.2) hovorí o Data Miningu ako o “Interdisciplinárnom poli spájajúcom techniky Machine Learning, rozoznávania vzorov, štatistiky, databáz a vizualizácie s cieľom získať prakticky využiteľné informácie z veľkých dátových zdrojov.“ Data mining môžeme teda charakterizovať ako proces analýzy dát s úmyslom nájsť skryté vzory a súvislosti v nich s použitím softvéru a v ňom integrovaných štatistických postupov (algoritmov). Všetko sa to deje za pomoci výpočtovej techniky, kde platí priama závislosť medzi objemom dát, podrobnosťou analýzy a potrebným výkonom hardvérovej platformy.

Data mining je významnou súčasťou Business Intelligence a preto je vhodné ozrejmiť niektoré súvislosti. Business Intelligence prvýkrát definoval Howard Dresner zo spoločnosti Gartner Group v roku 1989, ako „množinu konceptov a metodík, ktoré zlepšujú rozhodovací proces za použitia metrik alebo systémov založených na metrikách.“

Aj proces Data Miningu, ako každý iný proces, má istú štandardizovanú postupnosť krokov. Účelom takejto štandardizácie je spoľahlivosť a opakovateľnosť tohto procesu. Najpoužívanejšia data miningová metodológia opisujúca efektívny postup pri realizácii data miningového riešenia je metodológia **CRoss Industry Standart Process for data mining** v skratke CRISP. Pomáha vyhnúť sa opakovaným chybám. Keďže sa vyznačuje univerzálnosťou, je nezávislá od typu úlohy, ktorou sa data mining zaoberá.

### Úlohy pre Data Mining

Techniky Data Miningu môžu byť aplikované prakticky vo všetkých sférach. Odpovedajú na rôzne typy otázok a viažu sa s nimi rôzne typy úloh. Pre každú otázku, ktorú chceme Data Miningom zodpovedať, existuje viacero metód, čiže typov úlohy. V niektorých prípadoch si vystačíme s jedným typom úlohy, v iných budeme musieť skombinovať viacero typov úloh, aby sme dosiahli výsledok.

**Klasifikácia** je jednou z bežných úloh. Spočíva v priradení kategórie každému skúmanému prípadu. Obvykle sa vyznačia atribúty ako vstupy, na základe ktorých bude prebiehať priradenie kategórií. Typickým príkladom algoritmov, ktoré sa využívajú na klasifikáciu, sú: Rozhodovacie stromy, Neurónové siete a Naive Bayesove algoritmy.

**Asociácie** alebo ich variant v podobe analýzy nákupného koša, teda identifikácia produktov, ktoré sa často predávajú spoločne. Cieľom je odhalenie asociačných pravidiel (ak zákazník kúpi A a B, potom s určitou pravdepodobnosťou kúpi aj C).

**Regresia** je podobná klasifikácii s výnimkou toho, že nehľadá vzory na opísanie triedy pre klasifikáciu, ale jej cieľom je nájsť vzory pre určenie numerickej hodnoty. Dokonalejšie algoritmy dokážu spracovať aj kategorické vstupy, nielen číselné. Regresia rieši problém typu: predikcia rýchlosti vetra na základe teploty, tlaku vzduchu a vlhkosti (MacLennan, 2009, s.8).

**Predikcia** je jednou z najzaujímavejších možností Data Miningových analýz. Využíva sa na predpovedanie nových prípadov existujúceho radu hodnôt. Môže ísť o hodnotu cien aktív na burze, počet zákazníkov alebo iné.

Spomenuli sme len základné typy analýz. V praxi sa stretne s ďalšími typmi Data Miningových analýz, napr. Sekvenčnou analýzou, ktorá nachádza vzory v sériách udalosti tzv. „sekvenciách“ (MacLennan, 2009, s.9).

## 2. Akademický Informačný Systém - AiS2

Akademický informačný systém je komplexný informačný systém určený predovšetkým na riadenie všetkých stupňov vysokoškolského štúdia verejných vysokých škôl. Treba podotknúť, že v súčasnosti sa na jeho označenie využíva skratka AiS2, aby sa zreteľne odlíšilo, že ide o druhú verziu tohto systému.

Tento informačný systém bol pôvodne vyvinutý na Prírodovedeckej fakulte UPJŠ v Košiciach v súvislosti s prechodom na nový systém štúdia, kompatibilný s Európskym systémom prenosu kreditov (ECTS). Cieľom zavedenia AiS bolo i celkove zefektívnenie riadenia štúdia a fakulty ako celku. AiS môže byť prevádzkovaný na univerzite ako celku alebo jednotlivých fakultách. AiS2 je informačný systém s klasickou trojvrstvovou architektúrou. Na prístup do AiS2 sa používa štandardný internetový prehliadač, čo výrazne znižuje náklady na údržbu systému. Používateľské rozhranie je optimalizované pre prehliadače Internet Explorer a Mozilla Firefox.

Je dôležité spomenúť, že na projekte vývoja AiS2 participuje firma ORACLE Slovensko, s.r.o. Bratislava. Táto spoločnosť poskytla špeciálne zvýhodnené licencie pre svoje produkty. Spomínaná informácia je pre nás dôležitá aj z hľadiska databázového zdroja. Keďže na prevádzku a vývoj sú poskytnuté technické prostriedky a softvér od spoločnosti ORACLE, je tomu prispôbený aj súborový systém databázy. Pre účely skúmania dát prostredníctvom nástrojov Business Intelligence v MS SQL Serveri bolo potrebné realizovať konverziu dát z dôvodu rozdielnosti týchto databázových systémov.

### **Príprava databázy**

Databázový extrakt z AiS2 disponuje rozsiahlou štruktúrou s viac ako 700 tabuľkami. Konverzia dát z databázového systému ORACLE do databázového systému MS SQL Server bola zrealizovaná prostredníctvom nástroja MS SQL Server Migration Assistant for Oracle v januári 2011 a prebehla s niekoľkými komplikáciami:

- ⇒ nepreniesli sa všetky pôvodné tabuľky,
- ⇒ nepreniesli sa relačné vzťahy medzi tabuľkami,
- ⇒ dáta vo vybraných tabuľkách neboli k dispozícii v kompletnej forme, resp. neboli dostupné vôbec, tabuľky boli prázdne (tu by sa dalo polemizovať, či v pôvodnej databáze sú k dispozícii alebo ide o evidenčné nedostatky).

Prvý problém, keďže išlo len o malý počet tabuliek, s veľmi nízkou prioritou v rámci databázovej štruktúry, prináša len malé obmedzenie. Oveľa závažnejším nedostatkom bola absencia relačných vzťahov medzi tabuľkami. Našťastie existovala aspoň čiastočná dokumentácia schém, „relácií“ medzi tabuľkami, vďaka ktorým sa nám, po časovo náročnom úsilí, podarilo vytvoriť väčšinu relačných štruktúr.

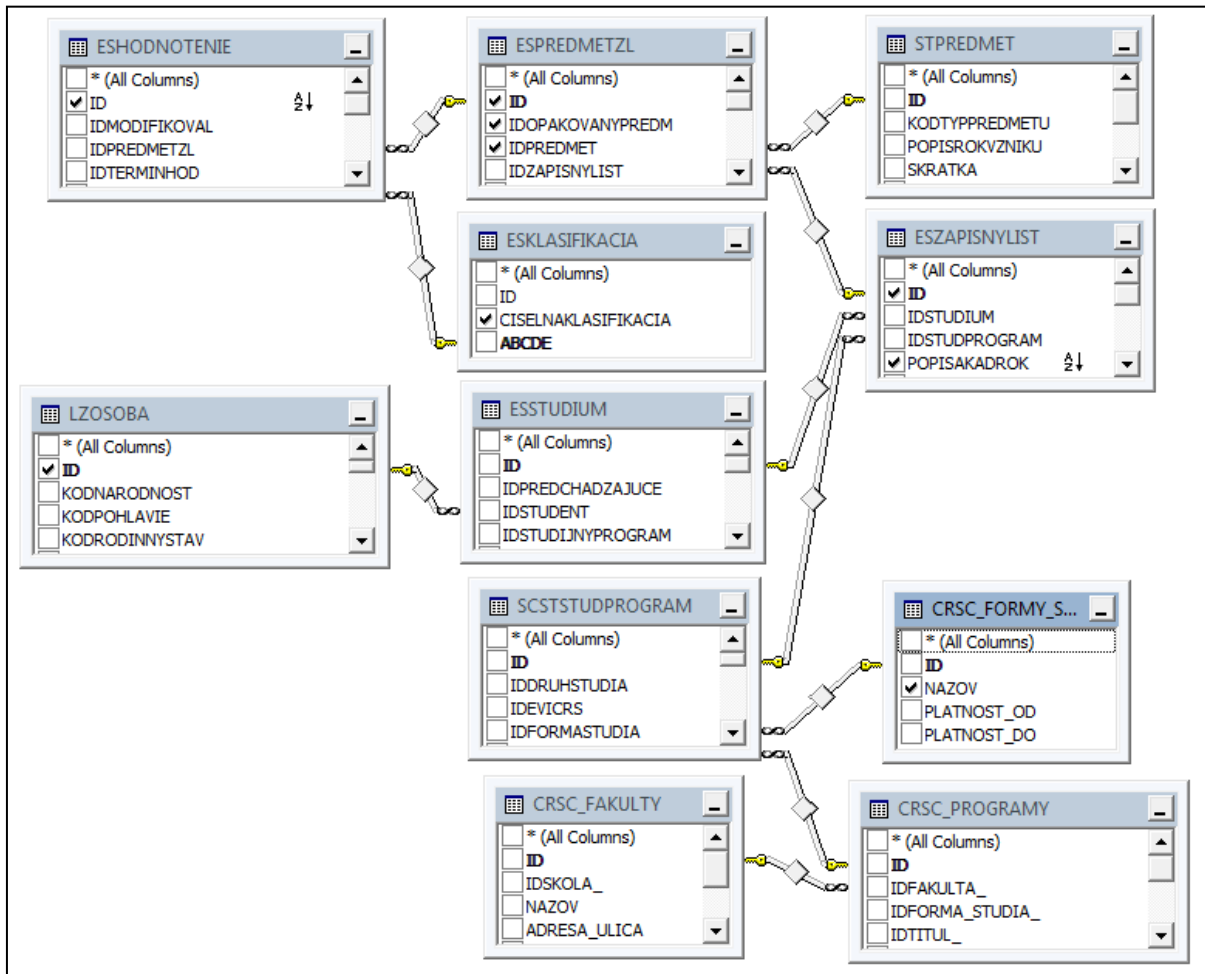
## **3. Analýza výsledkov hodnotení vybraných predmetov**

### **Príprava dát pre potreby analýzy**

Východiskovým zdrojom dát pre naše analýzy bola databáza vo formáte MDF, s veľkosťou približne 5,6 GB, prevádzkovaná v databázovom systéme MS SQL Server. Za účelom analýzy výsledkov hodnotení vybraných predmetov na Ekonomickej fakulte UMB (EF UMB) z jednotlivých termínov (riadny, opravný), sme získali dáta zahŕňajúce potrebné informácie o hodnoteniach študentov. Štruktúru tabuliek a ich relácií, potrebnú pre získanie dát o týchto hodnoteniach znázorňuje Obrázok 2.

Kľúčovou tabuľkou, z pohľadu prepojenia s ďalšími údajmi o študentoch, je LZOSOBA. Údaje o hodnoteniach študentov obsahuje tabuľka ESHODNOTENIE. Na jej prepojenie k našej kľúčovej tabuľke je potrebných ďalších 5 relačných vzťahov. Ostatné relácie sú súčasťou štruktúry z dôvodu potreby filtrovania údajov týkajúcich sa len EF UMB.

Obrázok 1: Schéma relačných vzťahov v prostredí Query Designera (Hodnotenia)



Zdroj: Vlastné spracovanie v prostredí Management Studio

Na začiatku sme určili, ktoré predmety podrobíme data miningovej analýze. Zvolili sme predmety s najvyšším počtom hodnotení v akad. roku 2009/2010 – Tabuľka 1.

Tabuľka 1: Analýza počtu hodnotení vo vybraných predmetoch

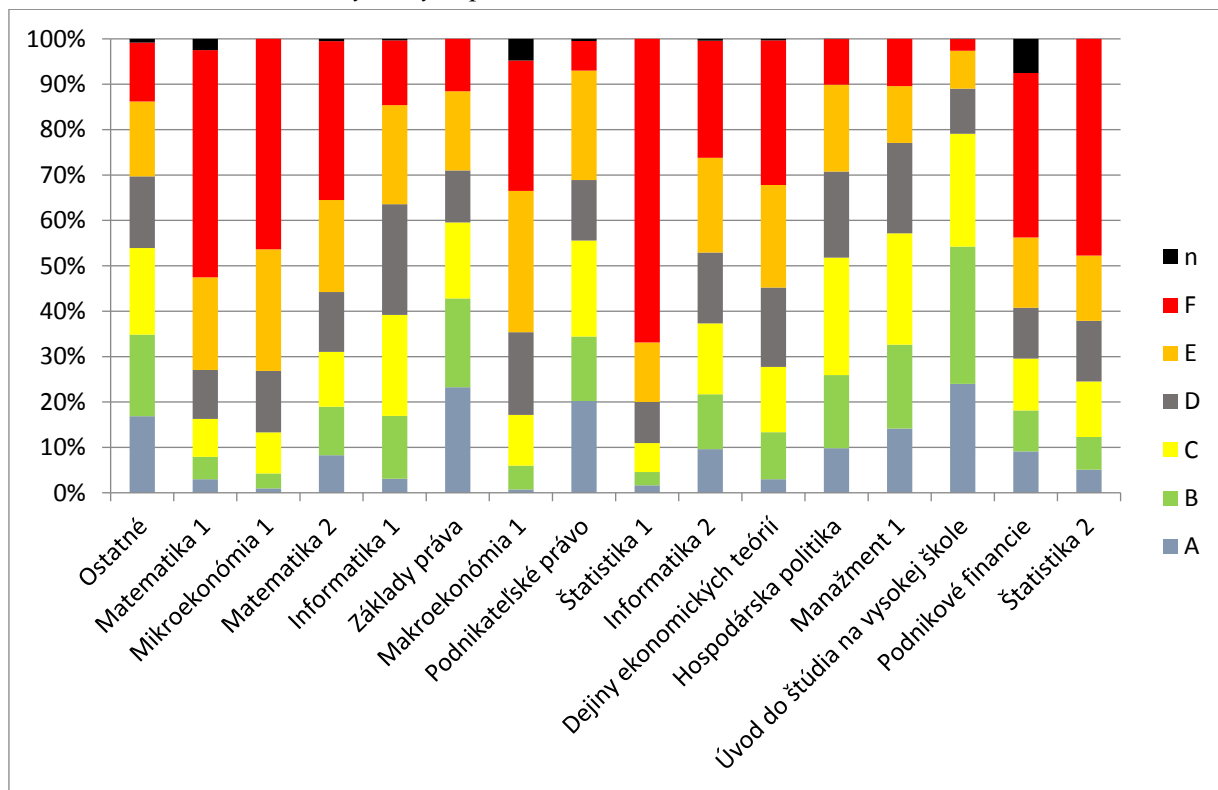
Počet hodnotení Riadny termín	Poradie	Názov Predmetu	Počet Hodnotení
	1	Matematika 1	649
	2	Mikroekonómia 1	617
	3	Matematika 2	551
	4	Informatika 1	544
	5	Základy práva	535
Počet hodnotení "A" Riadny termín	Poradie	Názov Predmetu	Počet Hodnotení
	1	Základy práva	124
	2	Podnikateľské právo	103
	3	Úvod do štúdia na vysokej škole	76
	4	Marketing	64
	5	Sociológia - RRV, VES (PP)	58

Počet hodnotení Opravný termín	Poradie	Názov Predmetu	Počet Hodnotení
	1	Štatistika 1	335
	2	Matematika 1	287
	3	Mikroekonómia 1	266
	4	Matematika 2	161
	5	Makroekonómia 1	159

Zdroj: Vlastné spracovanie

Ďalej sme skúmali štruktúru hodnotení najpočetnejších predmetov z riadneho termínu za sledované obdobie. Vertikálna os Grafu 1 znázorňuje proporčné množstvo hodnotení konkrétnou známku, označenú zodpovedajúcou farbou.

Graf 1: Štruktúra hodnotení vybraných predmetov



Zdroj: Vlastné spracovanie

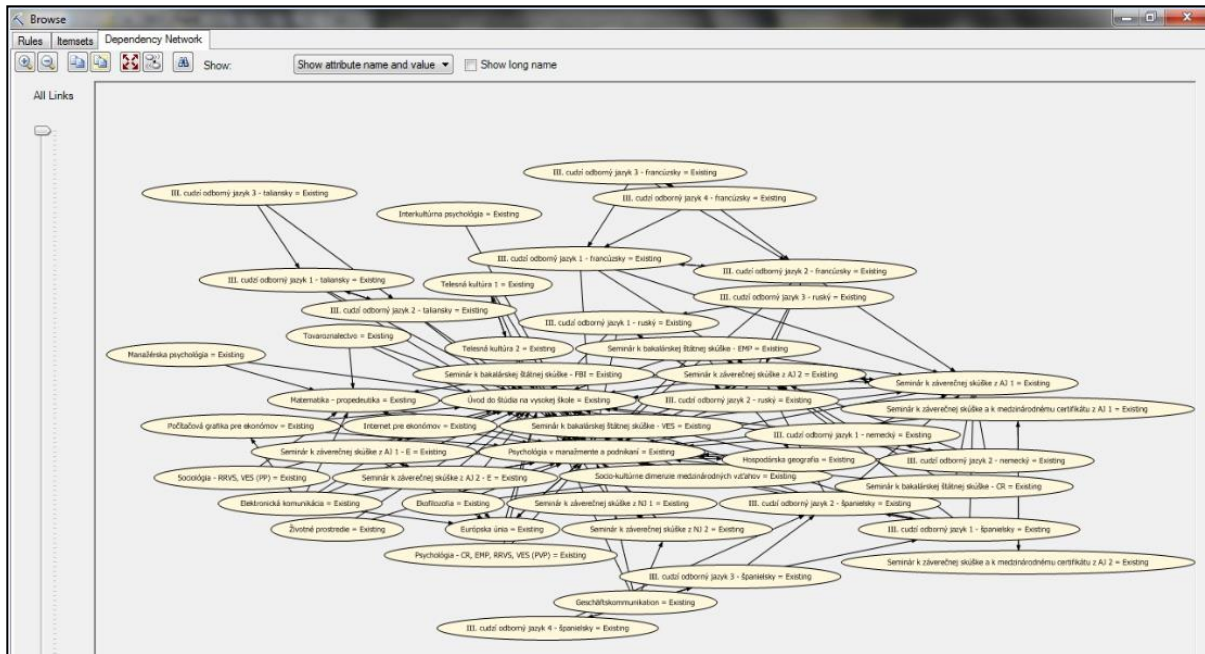
Na základe zobrazenej štruktúry môžeme označiť predmet Štatistika 1, ako predmet s najčastejším hodnotením Fx (viac ako 60%). Mohli by sme tu vidieť priestor na skúmanie faktorov, ktoré ovplyvňujú, že študent dosiahol spomínané hodnotenie. Jeho protipólom je predmet Úvod do štúdia na vysokej škole - proporčne obsahuje najväčšie zastúpenie hodnotenia „A“ z uvedených predmetov (nad 20%).

Veľmi známou analýzou v súvislosti s data miningom je analýza nákupného koša. Snaží sa zistiť, aké kombinácie produktov sa vyskytujú súčasne ako položky v jednom nákupe. Skúma sa pravdepodobnosť a početnosť ich výskytu. Analogicky môžeme tento typ analýzy uskutočniť aj na dátach z AiS2 databázy. Obdobou nákupného koša by mohol byť zápisný list a konkrétnou položkou v koši bude predmet. Výsledkom takéhoto skúmania by boli asociácie len v rámci jednotlivých akademických rokov. Možnosť výberu jednotlivých výberových predmetov presahuje akademický rok. Existuje možnosť vybrať si výberový predmet v prvom ročníku štúdia alebo aj v druhom, prípadne v treťom. Po zvážení, budeme ako

o nákupnom koši v kontexte analýzy nákupného koša uvažovať nie so zápisným listom, ale s identifikátorom osoby. Výsledkom toho bude hlbšia možnosť skúmania asociácií aj naprieč rokmi štúdia.

Obrázok 3 ukazuje jeden z troch výstupov algoritmu Microsoft Association Rules v prostredí Mining Viewers. V tomto prostredí poskytuje možnosť editácie zobrazenia stupňa najsilnejších asociácií. Na obrázku je znázornený stav asociácií pri nastavení „all links“ a teda sú zobrazené všetky spojenia.

Obrázok 2: *Dependency Network Viewer* výstupy



Zdroj: Vlastné spracovanie

V skúmaných dátach bolo vybraných 233 predmetov s tým, že niektoré z nich boli vo variante pre externistov. V Tabuľke 2 sa nachádza zoznam najčastejšie volených výberových predmetov:

Tabuľka 2: *Association Rules Viewer* výstupy 1

Poradie	Support	Size	Itemset
1	1457	1	Psychológia v manažmente a podnikaní = Existing
2	1421	1	Úvod do štúdia na vysokej škole = Existing
3	831	1	Matematika - propedeutika = Existing
4	731	1	Hospodárska geografia = Existing
5	666	1	Tovaroznalectvo = Existing
6	648	1	Seminár k záverečnej skúške z AJ 1 = Existing
7	601	1	Seminár k záverečnej skúške z AJ 2 = Existing
8	496	1	Európska únia = Existing
9	486	1	Internet pre ekonómov = Existing
10	392	1	III. cudzí odborný jazyk 1 - španielsky = Existing

Zdroj: Vlastné spracovanie

Tabuľka 3 ponúka pohľad na najzastúpenejšie dvojice predmetov a analogicky Tabuľka 4 zase zahŕňa najzastúpenejšie trojice.

Tabuľka 3: Association Rules Viewer výstupy 2

Poradie	Support	Size	Itemset
1	926	2	Úvod do štúdia na vysokej škole = Existing, Psychológia v manažmente a podnikaní = Existing
2	557	2	Seminár k záverečnej skúške z AJ 2 = Existing, Seminár k záverečnej skúške z AJ 1 = Existing
3	530	2	Hospodárska geografia = Existing, Psychológia v manažmente a podnikaní = Existing
4	523	2	Matematika - propedeutika = Existing, Psychológia v manažmente a podnikaní = Existing
5	437	2	Tovaroznactvo = Existing, Úvod do štúdia na vysokej škole = Existing
6	420	2	Seminár k záverečnej skúške z AJ 1 = Existing, Psychológia v manažmente a podnikaní = Existing
7	400	2	Matematika - propedeutika = Existing, Úvod do štúdia na vysokej škole = Existing
8	397	2	Hospodárska geografia = Existing, Úvod do štúdia na vysokej škole = Existing
9	380	2	Seminár k záverečnej skúške z AJ 2 = Existing, Psychológia v manažmente a podnikaní = Existing
10	363	2	Seminár k záverečnej skúške z AJ 1 = Existing, Úvod do štúdia na vysokej škole = Existing

Zdroj: Vlastné spracovanie

Tabuľka 4: Association Rules Viewer výstupy 3

Poradie	Support	Size	Itemset
1	357	3	Seminár k záverečnej skúške z AJ 2 = Existing, Seminár k záverečnej skúške z AJ 1 = Existing, Psychológia v manažmente a podnikaní = Existing
2	310	3	Seminár k záverečnej skúške z AJ 2 = Existing, Seminár k záverečnej skúške z AJ 1 = Existing, Úvod do štúdia na vysokej škole = Existing
3	307	3	Hospodárska geografia = Existing, Úvod do štúdia na vysokej škole = Existing, Psychológia v manažmente a podnikaní = Existing
4	264	3	Seminár k záverečnej skúške z AJ 1 = Existing, Úvod do štúdia na vysokej škole = Existing, Psychológia v manažmente a podnikaní = Existing
5	247	3	Matematika - propedeutika = Existing, Úvod do štúdia na vysokej škole = Existing, Psychológia v manažmente a podnikaní = Existing
6	244	3	Tovaroznactvo = Existing, Úvod do štúdia na vysokej škole = Existing, Psychológia v manažmente a podnikaní = Existing
7	241	3	Seminár k záverečnej skúške z AJ 2 = Existing, Úvod do štúdia na vysokej škole = Existing, Psychológia v manažmente a podnikaní = Existing
8	226	3	Internet pre ekonómov = Existing, Úvod do štúdia na vysokej škole = Existing, Psychológia v manažmente a podnikaní = Existing
9	225	3	Európska únia = Existing, Úvod do štúdia na vysokej škole = Existing, Psychológia v manažmente a podnikaní = Existing
10	198	3	Seminár k záverečnej skúške z AJ 2 = Existing, Seminár k záverečnej skúške z AJ 1 = Existing, Matematika - propedeutika = Existing

Zdroj: Vlastné spracovanie

V Tabuľke 5 sú záznamy poskytnuté výstupom asociačných pravidiel v časti „rules“. Ide o najsilnejšie asociácie medzi výberovými predmetmi. Väčšinou časť z nich tvoria asociácie medzi nadväzujúcimi predmetmi jazykov.

Tabuľka 5: Association Rules Viewer výstupy 4

Poradie	Probability	Importance	Rule
1	100 %	2,14	III. cudzí odborný jazyk 4 - španielsky = Existing -> III. cudzí odborný jazyk 3 - španielsky = Existing
2	99 %	1,46	III. cudzí odborný jazyk 2 - španielsky = Existing -> III. cudzí odborný jazyk 1 - španielsky = Existing
3	99 %	1,74	III. cudzí odborný jazyk 2 - francúzsky = Existing -> III. cudzí odborný jazyk 1 - francúzsky = Existing
4	99 %	1,73	III. cudzí odborný jazyk 2 - taliansky = Existing -> III. cudzí odborný jazyk 1 - taliansky = Existing
5	99 %	1,49	III. cudzí odborný jazyk 2 - ruský = Existing -> III. cudzí odborný jazyk 1 - ruský = Existing
6	98 %	1,06	III. cudzí odborný jazyk 3 - ruský = Existing -> III. cudzí odborný jazyk 2 - ruský = Existing
7	98 %	0,94	III. cudzí odborný jazyk 3 - ruský = Existing -> III. cudzí odborný jazyk 1 - ruský = Existing
8	97 %	2,35	III. cudzí odborný jazyk 4 - francúzsky = Existing -> III. cudzí odborný jazyk 3 - francúzsky = Existing
9	96 %	1,15	III. cudzí odborný jazyk 3 - taliansky = Existing -> III. cudzí odborný jazyk 1 - taliansky = Existing
10	96 %	1,87	Seminár k záverečnej skúške a k medzinárodnému certifikátu z AJ 2 = Existing -> Seminár k záverečnej skúške a k medzinárodnému certifikátu z AJ 1 = Existing

Zdroj: Vlastné spracovanie

Stĺpec „Probability“ hovorí o pravdepodobnosti výskytu kombinácie predmetov. Pravidlo je znázornené systémom, ktorý si ukážeme na prvom riadku. Interpretuje sa nasledovne: „Ak má študent v zápisnom liste predmet III. Cudzí odborný jazyk 2 - ruský s pravdepodobnosťou 100% bude mať v zápisnom liste aj predmet III. Cudzí odborný jazyk 1 – ruský“. Stĺpec Importance je interným skóre použitého algoritmu Association Rules a definovaný vzorcom:

$$\text{Importance} (\{A,B\}) = \text{Probability} (A, B) / (\text{Probability} (A) * \text{Probability} (B))$$

Vyjadruje akúsi dôležitosť danej asociácie. Čím vyššia je hodnota ukazovateľa importance, tým kvalitnejšie je pravidlo (MacLennan, 2009, s.358).

Extra analyzovaný výsledok dvojíc predmetov s odfiltrovanými predmetmi z jazykov (kde sú súvislosti zrejmé), je v Tabuľke 6. Prezentované záznamy sú výsekmi získaných výstupov.

Tabuľka 6: Association Rules Viewer výstupy 5

Poradie	Probability	Importance	Rule
1	70 %	0,11	Hospodárska geografia = Existing -> Psychológia v manažmente a podnikaní = Existing
2	68 %	0,06	Manažérska psychológia = Existing -> Úvod do štúdia na vysokej škole = Existing
3	67 %	0,06	Elektronická komunikácia = Existing -> Psychológia v manažmente a podnikaní = Existing
4	67 %	0,54	Elektronická komunikácia = Existing -> Internet pre ekonómov = Existing
5	67 %	0,64	Počítačová grafika pre ekonómov = Existing -> Internet pre ekonómov = Existing
6	67 %	0,06	Životné prostredie = Existing -> Úvod do štúdia na vysokej škole = Existing
7	64 %	0,04	Počítačová grafika pre ekonómov = Existing -> Úvod do štúdia na vysokej škole = Existing
8	64 %	0,05	Internet pre ekonómov = Existing -> Úvod do štúdia na vysokej škole = Existing
9	57 %	-0,02	Matematika - propedeutika = Existing -> Psychológia v manažmente a podnikaní = Existing
10	40 %	0,10	Internet pre ekonómov = Existing -> Matematika - propedeutika = Existing

Zdroj : Vlastné spracovanie

Dospeli sme k množstvu informácií, ktoré by boli lepšie využiteľné priamo v akademickom informačnom systéme. Ak opäť použijeme analógiu s analýzou nákupného koša, tak zistené výsledky by mohli slúžiť ako odporúčania pre kupujúcich, akú položku tovaru kombinovať s tým, ktorý už máme v nákupnom koši. Pre obchodný reťazec to znamená prispôbiť tovarové položky s vysokou mierou asociácie alebo korelácie, aj v súvislosti s umiestnením v obchode. Takéto korelované položky je potom vhodné umiestňovať v rámci nákupných priestorov blízko pri sebe. A teda odporúčané kombinácie predmetov by mohli byť za istých okolností zobrazované pri zápisných listoch v prostredí AiS2 spoločne.

## 4. Záver

Súčasný stav používania analytických aplikácií na Ekonomickej fakulte UMB dáva výrazný priestor na ich zlepšenie. Prezentované modely sú iba zlomkom možností, ako aplikovať technológiu Data Mining v oblasti riadenia verejnej vysokej školy.

Úlohou zodpovedných pracovníkov je posúdiť možnosti výberu najvhodnejšieho analytického systému, pripraviť a zrealizovať proces implementácie a vytvoriť predpoklady pre efektívne používanie týchto nástrojov v pravidelnom procese rozhodovania.



## 5. Literatúra

**AiS2 Informačný portál.** Zoznam Vysokých škôl využívajúcich systém AiS2. [online]. Dostupné na internete: <<http://www.ais2.sk/xwiki/bin/view/Referencie/>>

**Aitchison, A. 2009.** Expert SQL Server 2008 Development Advanced SQL Server techniques for database professionals. USA: Apress, 2009. ISBN 978-1-4302-7212-0.

**HAN J., KAMBER M.: 2006.** Data Mining – Concepts and Techniques, Second Edition. USA: Elsevier Inc., 2006. ISBN 978-1-55860-901-3.

**Langit, L., 2009.** Smart Business Intelligence Solutions with Microsoft SQL Server 2008. USA: Microsoft Press, 2009. ISBN 978-0-735-62580-8.

**LAROSE, D.T.: 2005.** Discovering Knowledge in Data – Introduction to Data Mining.

**Larson, B. 2009.** Delivering Business Intelligence with Microsoft SQL Server 2008. USA: McGraw-Hill Companies, 2009. ISBN 978-0-07-154945-5.

**MacLennan, J. 2009.** Data Mining with Microsoft SQL Server 2008. USA: John Wiley & Sons, Inc., 2009. ISBN 978-0-470-27774-4.

**PIZZUTI, C., RITCHIE, M. 2009.** Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics. Nemecko: Springer Berlin Heidelberg, 2009. ISBN 978-3-642-01183-2.

**Portál Adastra.cz.** Schéma Business Intelligence. [online]. Dostupné na internete: <<http://www.adastra.cz/dokument.aspx?id=192&archive=2007>>

**Portál batchgeo.com.** Nástroj pre konverziu údajov s adresami na Spatial data. [online]. Dostupné na internete: <<http://batchgeo.com/>>

**Portál crisp-dm.org.** Data Miningový cyklus CRISP. [online]. Dostupné na internete: <<http://www.crisp-dm.org/Process/index.htm>>

**Portál data-mine.com.** Overený zoznam aktuálnych data mining programov. [online]. Dostupné na internete: <<http://www.the-data-mine.com/bin/view/Software/AllDataMiningSoftware>>

**Portál easydatamining.com.** Príklady data miningového využitia. [online]. Dostupné na internete: <[http://www.easydatamining.com/index.php?option=com\\_content&view=category&layout=blog&id=29&Itemid=89&lang=en](http://www.easydatamining.com/index.php?option=com_content&view=category&layout=blog&id=29&Itemid=89&lang=en)>

**Portál kdnuggets.com.** Príklady data miningového využitia. [online]. Dostupné na internete: <<http://www.kdnuggets.com/solutions/index.html>>

**Portál Microsoft.com.** Limity a špecifikácie programu excel. [online]. Dostupné na internete: <<http://office.microsoft.com/en-us/excel-help/excel-specifications-and-limits-HP010073849.aspx>>

**Portál Univerzity P.J.Šafárika.** Charakteristika akademického informačného systému. [online]. Dostupné na internete: <<http://www.upjs.sk/filozoficka-fakulta/fakulta/ais/charakteristika/>>

USA: John Wiley & Sons, Inc., 2005. ISBN 0-471-66657-2.

## O autorovi

Jolana Gubalová sa v oblasti výskumu orientuje na databázové systémy a analýzu údajov z veľkých databáz s využitím technológií Business Intelligence a Data Mining. Zaujíma sa aj o oblasť Big Data a Internet of Things. Vyučuje predmety zamerané na manažérske informačné systémy a ERP systémy.