

**OBJAVENIE MEDZINÁRODNEJ JAZYKOVEJ AGENTÚRY
Z HĽADISKA DATAMINGU**

**DISCOVERY OF AN INTERNATIONAL LANGUAGE
AGENCY FROM THE POINT OF DATA MINING**

Mária Bohdalová – Eva Kurdyová

Abstract

Popularity of the data mining was reached when the companies became more interested in the needs of their customers and realized that if they want to gain competitive advantage, it is necessary to know the behaviour of their customers. Data mining is a process of data analysing in order to find something significant, something hidden. Among the best known data mining models is mentioned the technique of cluster analysis, which divides the sample into homogeneous units into clusters. The cluster analysis is used for segmentation in marketing. From a business perspective, each entrepreneur and manager is interested in the following questions: “Who is our target market?”, “What are the characteristics of our customers?”, “What will be the trends in the future and how many products can I sell in the next few months?”, “Have I enough suppliers?” or “Which product has the most powerful demand and Which business activity should be enhanced?”. In this paper with help of statistical and data mining methods we try to find the answers for these questions. Sample of purchase orders and clients provided by an international language agency were analysed by MS Excel 2010 and by statistical software IBM SPSS v. 20.

Abstrakt

Data mining dosiahol svoju popularitu na začiatku 21. storočia, keď sa spoločnosti začali zaujímať o potreby svojich zákazníkov a uvedomili si, že pre získanie konkurenčnej výhody je nevyhnutné poznať správanie sa zákazníka. Dataminig zahŕňa tie metódy business intelligence, ktorých cieľom je extrahovať významné informácie a znalosti z dát. Tieto metódy zahŕňajú napríklad matematicko – štatistické prístupy pre rozpoznávanie homogénnych skupín, ktoré sa používajú pre segmentáciu zákazníkov v marketingu. K najznámejším takým analýzám patrí zhluková analýza. Z podnikateľského hľadiska každého obchodníka a manažéra zaujímajú otázky: „Kto je naša cieľová skupina?“, „Akými charakteristikami disponujú naši zákazníci?“, „Aké budú trendy v budúcnosti a koľko výrobkov môžem predávať v nasledujúcich mesiacoch?“, „Mám dostatok dodávateľov?“

Reprodukce lidského kapitálu – vzájemné vazby a souvislosti. 9. – 10. prosince 2013

alebo „Ktorý produkt má najsilnejší dopyt a ktorú podnikateľskú aktivitu mám posilniť?“. Cieľom tohto príspevku je pomocou štatistických a dataminingových metód nájsť odpovede na tieto otázky pomocou tabuľkového procesora MS Excel 2010 a štatistického softvéru IBM SPSS v. 20. V príspevku sme analyzovali vzorku objednávok a zákazníkov, ktorú nám poskytla nemenovaná medzinárodná jazyková agentúra.

Key words: business intelligence, data mining, cluster analysis, time series analysis

Kľúčové slová: business inteligence, data mining, zhuková analýza, analýza časových radov

JEL Code: C12, C38, C81

Úvod

„Dobre riadiť firmu znamená dobre riadiť jej budúcnosť; pre dobré riadenie jej budúcnosti treba mať dobré informácie“ sú slová Mariana Harpera, vodcu v oblasti reklamy, v minulom storočí. Množstvo dát, ktoré máme k dispozícii má neustále rastúcu tendenciu, ale musíme si uvedomiť, že nie všetky dáta je možné pretransformovať do informácie. Ako príklad by sme uviedli, že ak berieme do úvahy všetky údaje, ktoré sú k dispozícii vo firmách a zo snímania v supermarketoch, má manažér v odbore baleného tovaru k dispozícii cca milión až miliardu nových údajov. Napriek vysokému počtu informácií manažéri a marketingoví pracovníci sa často sťažujú na *nedostatok informácií správneho druhu* alebo na *nadbytok nevhodných informácií*.

Dobre fungujúce malé a stredné podniky majú takzvaný „*learning relationship*“ so svojimi zákazníkmi, čo znamená, že pozorujú svojich zákazníkov, zbierajú informácie o ich nákupnom správaní, učia sa o nich a tieto znalosti sa snažia využiť pre lepšiu obsluhu. Výsledkom celého procesu sú nielen šťastní a verní zákazníci, ale aj ziskové podniky. Veľké spoločnosti, ktoré majú tisíc až milión zákazníkov, si nemôžu dovoliť osobný vzťah s každým z nich. Musia sa naučiť pracovať s údajmi, ktoré vznikajú pri styku so zákazníkmi. Datamining alebo dolovanie dát, je súbor analytických techník, ktoré pomáhajú firmám zmeniť zákaznicke dáta na „*customer knowledge*“ (Berry, Linoff, 2004).

1 Objavenie medzinárodnej jazykovej agentúry

Nemenovaná jazyková agentúra, s ktorou sme spolupracovali je orientovaná na potreby svojich zákazníkov. O klientov a objednávokach si vedie evidenciu a snaží sa vybudovať čo najlepší vzťah nielen so svojimi zákazníkmi ale aj s dodávateľmi.

Jazyková agentúra ponúka svojim zákazníkom možnosť výberu prekladateľského editora pri objednávke. Zákazník sa môže rozhodnúť o tom, či preferuje takzvaný obyčajný preklad v editoroch ako MS Word, MS Excel, MS Power Point, alebo potrebuje CAT¹ (computer-aided translation/počítačom podporovaný preklad) preklad v editoroch Trados alebo Idiom. Ak predpokladáme, že zákazník si cielene vyberá editor v závislosti od odboru prekladu, tak firma môže získať konkurenčnú výhodu na trhu prekladov. Pre firmu to znamená, že musí posilniť aj portfólio svojich prekladateľov s potrebnými prekladateľskými editormi. Predpokladanú súvislosť medzi odborom prekladu a výberom editora podľa krajiny pôvodu (Česká republika, Slovenská republika, všetci zákazníci) sformulovanú do hypotézy H_0 sme overili pomocou χ^2 kvadrát testu (Pacáková, 2003):

H_0 : Zákazník sa náhodne rozhodne pre výber editora (medzi premennými „Odbor“ a „Editor“ nie je závislosť)

H_1 : Zákazník cielene vyberá prekladateľský editor (medzi premennými „Odbor“ a „Editor“ je závislosť)

Hypotézu sme overovali na hladine významnosti $\alpha = 0,05$.

Tab. 1: χ^2 test pre premenné „Odbor“ a „Editor“ podľa krajiny pôvodu

<i>Chi-Square Tests</i>				
<i>Zákazník sa pochádza z CZ alebo SK?</i>		<i>Value</i>	<i>df</i>	<i>Asymp. Sig. (2-sided)</i>
<i>Zákazník CZ</i>	Pearson Chi-Square	608225,636 ^b	420	0,000
<i>Zákazník SK</i>	Pearson Chi-Square	66358,551 ^c	380	0,000
<i>Všetci zákazníci</i>	Pearson Chi-Square	616065,182 ^a	430	0,000

Zdroj: spracovanie vlastné pomocou IBM SPSS

Z tabuľky 1 vidíme, že vypočítaná p -hodnota je menšia v porovnaní s nami uvažovanou hodnotou štatistickej významnosti, ktorá je 5%. H_0 , podľa ktorej sa zákazník náhodne rozhodne pri výbere prekladateľského editora, sa zamietá (p -hodnota = 0,000). Potvrdila sa teda alternatívna hypotéza, ktorá svedčí o závislosti medzi premennými „Odbor“ a „Editor“ v jednotlivých krajinách a aj v oboch krajinách.

χ^2 test nám potvrdil, že zákazníci (slovenskí aj českí) cielene vyberajú editor na základe odboru prekladu. Napríklad slovenskí zákazníci objednávajú medicínske preklady v editore Trados Studio (30,1%) alebo v Trados TagEditor-e (15,5%). Niektoré editory sa používajú

¹ CAT program spolupracuje s prekladovou pamäťou a terminologickou databázou. Pomocou týchto nástrojov sa znižuje chybovosť prekladu ale aj čas, potrebný na prekladanie.

RELIK 2013.

Reprodukce lidského kapitálu – vzájemné vazby a souvislosti. 9. – 10. prosince 2013

len pre jeden vedný odbor (pre stavebníctvo je to Adobe Photoshop 0,1% alebo AutoCAD 0,1%; pre informačné technológie sú to editory: Alchemy Catalist 0,8%, Idiom 0,5%, Star Transit 0,1%)

V tabuľke 2 uvádzame podielové zastúpenie jednotlivých editorov pre jednotlivé vedné odbory (výsledky vyplývajú zo vzorky objednávok). Dá sa povedať, že najpoužívanejší editor pre vyhotovenie prekladov prakticky pre všetky vedné odbory je MS Word.

Tab. 2 Asociačná tabuľka pre odbor prekladu a prekladateľský editor

	nezadané	ART	BUIL	ECO	HUM	CHEM	ITEL	LAW	MEDI	NATU	TECH
nezadané	96,70%	9,10%	6,00%	5,40%	4,20%	3,30%	4,30%	4,50%	12,70%	6,40%	3,40%
Across	0,10%	8,80%	5,50%	1,60%	0,40%	1,60%	1,50%	1,20%	4,20%	2,10%	2,40%
Adobe Acrobat	0,60%	4,20%	10,10%	13,30%	16,50%	4,20%	1,30%	10,60%	6,20%	6,80%	3,20%
Adobe FrameMaker	0,10%	0,20%	5,60%	0,50%	1,10%	1,60%	0,20%	0,10%	0,00%	0,10%	0,00%
Adobe Illustrator	0,20%	0,00%	1,00%	0,10%	0,50%	0,70%	0,10%	0,00%	0,00%	0,00%	0,00%
Adobe InDesign	0,30%	0,10%	0,60%	0,10%	0,50%	0,80%	0,10%	0,00%	0,10%	0,20%	0,10%
Adobe PageMaker	0,00%	0,00%	0,10%	0,00%	0,10%	0,10%	0,00%	0,00%	0,00%	0,00%	0,00%
Adobe Photoshop	0,10%	0,00%	0,10%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
Alchemy Catalyst	0,00%	0,00%	0,00%	0,10%	0,00%	0,00%	0,80%	0,00%	0,00%	0,00%	0,00%
AutoCAD	0,00%	0,00%	0,10%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
Idiom	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,50%	0,00%	0,00%	0,00%	0,00%
Logoport	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,60%	0,00%	0,40%	0,00%	0,10%
MS Access	0,00%	0,40%	0,30%	0,60%	0,50%	0,20%	0,10%	0,60%	0,30%	0,30%	0,10%
MS Excel	0,10%	2,70%	5,50%	2,20%	2,80%	14,00%	8,40%	0,30%	2,00%	4,20%	2,80%
MS LocStudio	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,60%	0,00%	0,00%	0,00%	0,00%
MS PowerPoint	0,00%	0,90%	0,50%	2,10%	0,60%	0,60%	0,30%	0,10%	0,90%	1,00%	0,80%
MS Visio	0,00%	0,10%	0,10%	0,10%	0,00%	0,00%	0,00%	0,10%	0,10%	0,10%	0,00%
MS Word	0,80%	58,50%	62,50%	68,40%	61,20%	52,70%	23,70%	70,20%	46,50%	72,90%	43,70%
NEMO (CAT tool)	0,00%	6,10%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	23,30%
Notepad	0,00%	0,10%	0,00%	0,00%	0,00%	0,00%	0,30%	0,00%	0,00%	0,00%	0,10%
OCR (Abbyy Fine Reader)	0,10%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
ostatné	0,10%	0,90%	0,60%	1,30%	1,80%	0,80%	1,40%	2,20%	0,70%	0,70%	4,70%
Passolo	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,20%	0,00%	0,00%	0,00%	0,10%
QuarkXPress	0,10%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
SDL Trados Studio	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	1,10%	0,00%	0,60%	0,00%	0,40%
Star Transit	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,10%	0,00%	0,00%	0,00%	0,10%
Trados Multiterm	0,00%	0,20%	0,10%	0,20%	0,20%	0,40%	4,20%	0,20%	1,00%	0,30%	1,50%
Trados Studio	0,00%	0,10%	0,00%	0,00%	0,10%	0,10%	0,90%	0,00%	0,20%	0,10%	0,70%
Trados TagEditor	0,10%	2,70%	0,30%	0,90%	2,20%	3,10%	35,80%	0,20%	9,30%	1,30%	6,90%
Trados WorkBench	0,50%	4,60%	0,80%	2,80%	6,70%	15,70%	13,00%	9,50%	14,70%	3,30%	5,50%
Wordfast	0,00%	0,30%	0,00%	0,20%	0,60%	0,00%	0,30%	0,00%	0,00%	0,30%	0,00%
XML editor	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,10%	0,10%	0,00%	0,00%	0,00%

Zdroj: spracovanie vlastné pomocou IBM SPSS

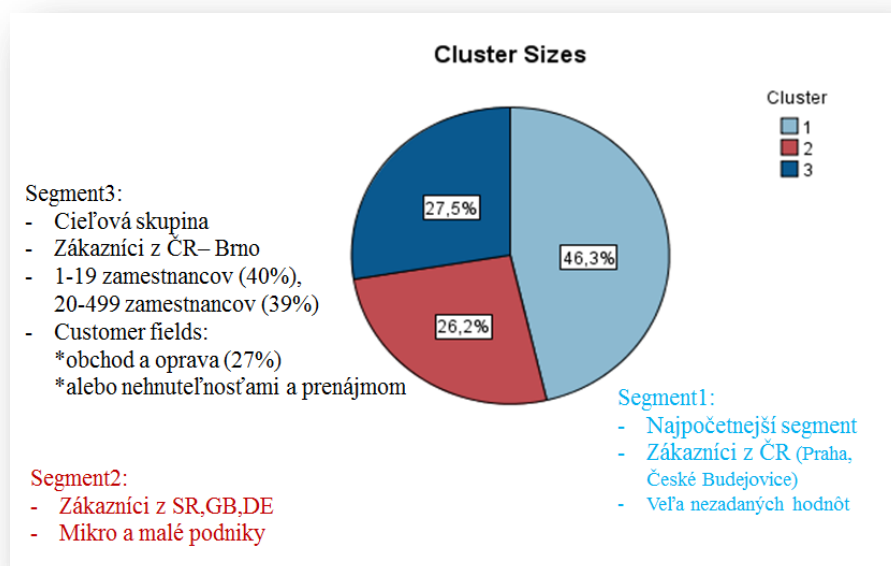
Z podnikateľského hľadiska každého obchodníka a manažera zaujímajú otázky: „Kto je naša cieľová skupina?“ a „Akými charakteristikami disponujú naši zákazníci?“. Odpovede na položené otázky získame pomocou vybraných metód data miningu. Analýza dát pomocou data miningu predstavuje iteratívny proces zameraný na analýzu veľkých databáz, ktorých účelom je získavať informácie a vedomosti, ktoré sa môžu javiť ako presné a potenciálne užitočné pre pracovníkov zaoberajúcich sa rozhodovaním a riešením problémov (Vercellis, 2009). Účelom data miningovej analýzy je vyvodiť určité závery zo vzorky skorších pozorovaní a zovšeobecniť tieto závery vzhľadom na celú populáciu, a to takým spôsobom,

Reprodukce lidského kapitálu – vzájemné vazby a souvislosti. 9. – 10. prosince 2013

že sú tak presné ako je to len možné . Modely a vzory zistené týmto spôsobom môžu mať rôzne formy a môžu byť získané rôznymi metódami. V tomto príspevku využijeme zhlukovú analýzu (Řezánková, 2009) pre identifikáciu segmentov cieľovej skupiny zákazníkov a analýzu časových radov (Rublíková, 2007), (Vercellis, 2009) pre predikovanie nákupného správania sa zákazníkov.

Cieľovú skupinu jazykovej agentúry sme hľadali pomocou štatistického programu IBM SPSS Statistics 20. Na základe demografických a nákupných charakteristík: *štát, mesto, sektor a počet zamestnancov firmy* sme identifikovali tri zákaznicke segmenty, ktoré sú znázornené na obrázku 1.

Obr. 1 Výsledky zhlukovej analýzy



Zdroj: spracovanie vlastné pomocou IBM SPSS

Segment 1: Prvý, najpočetnejší segment zahŕňa zákazníkov z Českej republiky, z mesta Praha, České Budějovice alebo Karlove Vary. Pre klientov prvého segmentu nebolo možné definovať sektor, v ktorom podnikajú z dôvodu nezadaných hodnôt v databáze. Ak neberieme do úvahy nezadané hodnoty pri počte zákazníkov firiem, môžeme vyhlásiť, že väčšina firiem zamestnáva 1–19 zamestnancov. V porovnaní s ostatnými segmentmi, zhluk č.1 obsahuje najviac nezadaných hodnôt.

Segment 2: Do zhluku patrí 27,5% zákazníkov firmy, ktorí cca s 60%–nou pravdepodobnosťou pochádzajú zo Slovenskej republiky. Segment v malom počte zahŕňa aj zákazníkov z Veľkej Británie a z Nemecka. 20,5% zákazníkov má svoje korene v Bratislave.

Reprodukce lidského kapitálu – vzájemné vazby a souvislosti. 9. – 10. prosince 2013

Firmy patria do kategórie mikro a malých podnikov, avšak oblasť ich podnikateľskej činnosti nie je možné zistiť.

Segment3: Ak sa nemenovaná jazyková agentúra rozhodne orientovať na tretí segment, jej cieľovou skupinou môžu byť zákazníci, ktorí pochádzajú z Českej republiky, z mesta Brno. Firmy v tom segmente s 40,4%–nou pravdepodobnosťou zamestnávajú 1–19 zamestnancov a približne s 39%–nou pravdepodobnosťou 20–499 zamestnancov. Najväčší počet klientov pracuje v oblasti obchodu a opravy (27,3%), alebo sa zaoberá nehnuteľnosťami a prenájmom. Ak nemenovaná jazyková agentúra chce získať konkurenčnú výhodu na trhu prekladateľských služieb musí vzhľadom na rastúci spracovateľský priemysel zabezpečiť spoluprácu s prekladateľmi z oblasti techniky a chémie, aby vedela uspokojiť potrebu firiem z tejto oblasti.

Podarilo sa nám nájsť cieľovú skupinu jazykovej agentúry, identifikovali sme charakteristiky, ktorými jednotlivé zákaznícke segmenty disponujú, odhalili sme skrytú závislosť medzi editorom a odborom prekladu. Zostali nám ďalšie dve otázky, ktoré do značnej miery ovplyvňujú chod firmy v budúcnosti. Otázky sú nasledovné: „Aké budú trendy v budúcnosti a koľko výrobkov môžeme predávať v nasledujúcich mesiacoch?“.

Trendy v nasledujúcich mesiacoch súvisia so vznikom nových firiem v blízkosti jednotlivých pobočiek jazykovej agentúry a od zákazníkov, s ktorými už firma spolupracuje. Pre odhad trendu v budúcnosti, budeme analyzovať doterajšiu situáciu firmy, zákazníkov, s ktorými spolupracuje a počet objednávok v posledných mesiacoch. Z databázy objednávok a zákazníkov jazykovej agentúry vyplýva, že 81% objednávok sa týka prekladu a 15,6% korektúr, pričom najčastejšie boli objednané nasledujúce jazykové kombinácie:

TRCSEN	- preklad z češtiny do angličtiny (45772krát, 9%)
TRENCS	- preklad z angličtiny do češtiny (27978krát, 5,5%)
TRCSDE	- preklad z češtiny do angličtiny (22607krát 4,4%)
PEEN	- jazyková korektúra anglického jazyka (16506krát, 3,2%)
TRDECS	- preklad z nemčiny do češtiny (15922krát, 3,1%)
TWENCS	- súdny preklad z angličtiny do češtiny (11846krát, 2,3%)

Najvyhľadávanejšie odbory boli: technika (19,2%), právo (16,2%), ekonomika a financie (12,7%), štylistické a literárne texty (11,5%) a medicína (5,3%). Väčšina prekladov bola objednaná za účelom publikovania (46,2%) alebo mali informatívny charakter (42,1%). Najpopulárnejšie editory pri prekladaní sú stále MS Word (43%), Adobe Acrobat (5,9%) a MS Excel (1,9%); z CAT nástrojov Trados WorkBench (5,3%), Trados TagEditor (3,7%) a Across (2,3%).

Zistili sme, že 8,2% odberateľov sa zaoberá *obchodom*, 7% zákazníkov sa zaoberá *nehnutelnosťami, prenájmom a službami*, 6,7% podniká v oblasti *spracovateľského priemyslu*. 65% zákazníkov podniká samostatne alebo nevykazuje podnikateľskú aktivitu, 15% zákazníkov tvoria firmy s počtom zamestnancov 1 až 19. 12,6% klientov zamestnáva 20 až 499 zamestnancov. 9101 zákazníkov sa rozhodlo pre jazykové služby firmy na základe odporúčenia a 7085 na základe vyhľadávania na internete.

Pre otázku, „Koľko výrobkov môžeme predávať v nasledujúcich mesiacoch“ sme analyzovali len preklady z anglického jazyka do slovenského jazyka. Dôvod bol ten, že ak berieme do úvahy všetky jazykové kombinácie pri preklade, pri jazykovej korektúre jedno a dvojjazyčnej a keď vezmeme do úvahy ďalšie služby, ktoré firma ponúka zákazníkom a každý jeden z nich uvažujeme ako jeden osobitný produkt, tak firma disponuje s viac ako 1500 produktmi.

Počet prekladov z anglického jazyka do slovenského jazyka sme analyzovali od apríla 2007 až do septembra 2012, čo znamená 66 po sebe idúcich hospodárskych mesiacov. Na obrázku 2 vidíme, že hospodárska kríza, ktorú pocítila Európa v prvom štvrtroku 2008, ovplyvnila aj počet objednaných prekladov z anglického jazyka do jazyka slovenského. Najviac prekladov tohto typu bolo objednaných v júli roku 2010, presnejšie 180, a najmenej v apríli a máji roku 2010, kedy zákazníci objednali 51 a 46 prekladov z anglického jazyka do slovenského jazyka.

Na prognózu počtu prekladov sme používali štatistické metódy, presnejšie časové rady (Rublíková, 2007), (Pacáková, 2003), (Vercellis, 2009). Pri výbere funkcie trendu sme vychádzali z nasledujúcich výberových kritérií:

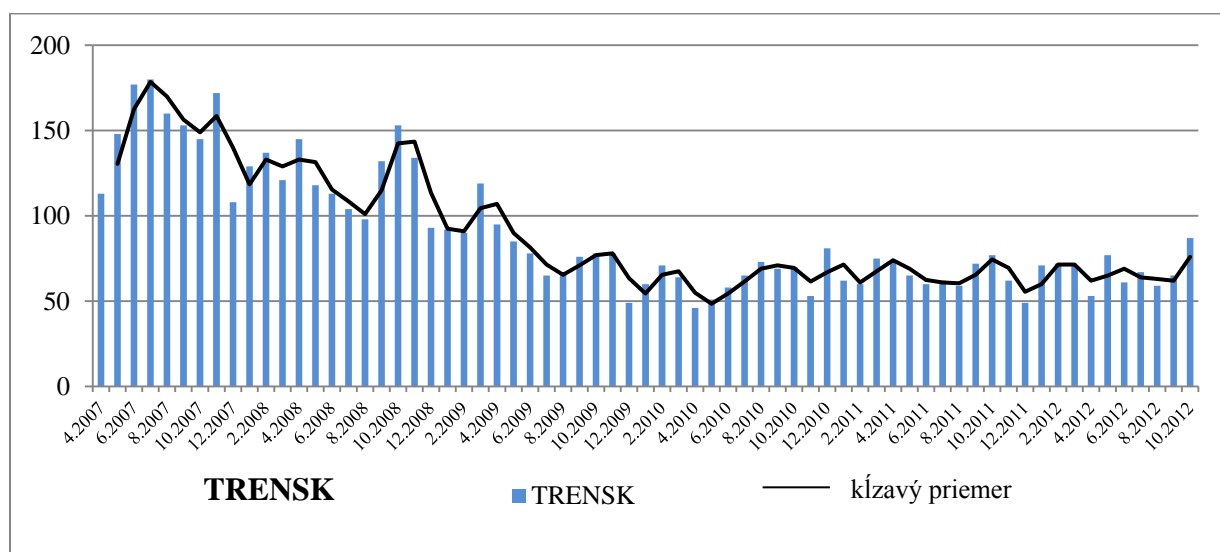
- a) Ak prvé absolútne diferencie $\Delta t = y_t - y_{t-1}$ v čase $t = 2, 3, \dots$ kolíšu okolo nuly, tak použijeme **konštantný trend**.
- b) Ak prvé absolútne diferencie $\Delta t = y_t - y_{t-1}$ pre čas $t = 2, 3, \dots$ kolíšu okolo nenulovej hodnoty (konštanty) a druhé absolútne diferencie $\Delta_t^2 = \Delta_t - \Delta_{t-1}$ pre čas $t = 3, 4, \dots$ kolíšu okolo nuly, tak použijeme **lineárny trend**.
- c) Ak prvé absolútne diferencie $\Delta t = y_t - y_{t-1}$ pre čas $t = 2, 3, \dots$ vykazujú lineárny trend, pritom druhé absolútne diferencie $\Delta_t^2 = \Delta_t - \Delta_{t-1}$ kolíšu okolo nenulovej hodnoty a tretie absolútne diferencie $\Delta_t^{(3)} = \Delta_t^{(2)} - \Delta_{t-1}^{(2)}$ spĺňajú kritérium, že v čase $t = 4, 5, \dots$ kolíšu okolo nuly, odporúča sa použiť **kvadratický trend**.
- d) Pre **exponenciálny trend** je typické, že koeficienty rastu $k_t = \frac{y_t}{y_{t-1}}$ pre čas $t = 2, 3, \dots$ sú približne konštantné.

Tab. 3 Výber analytickej funkcie trendu

Výber analytickej funkcie trendu						
Dátum	TRENSK		konštantný trend	lineárny model	kvadratický model	exponenciálny trend
	$y(t)$	t	1. absolútne diferencie $\Delta t = y_t - y_{t-1}$	2. absolútne diferencie $\Delta_t^{(2)} = \Delta_t - \Delta_{t-1}$	3. absolútne diferencie $\Delta_t^{(3)} = \Delta_t^{(2)} - \Delta_{t-1}^{(2)}$	koefficient rasru $k_t = \frac{y_t}{y_{t-1}}$
4.2007	113	1	-	-	-	-
5.2007	148	2	35	-	-	1,310
6.2007	177	3	29	-6	-	1,196
7.2007	180	4	3	-26	-20	1,017
8.2007	160	5	-20	-23	3	0,889
9.2007	153	6	-7	13	36	0,956
10.2007	145	7	-8	-1	-14	0,948
11.2007	172	8	27	35	36	1,186
12.2007	108	9	-64	-91	-126	0,628
1.2008	129	10	21	85	176	1,194
2.2008	137	11	8	-13	-98	1,062
3.2008	121	12	-16	-24	-11	0,883
4.2008	145	13	24	40	64	1,198
5.2008	118	14	-27	-51	-91	0,814
6.2008	113	15	-5	22	73	0,958
7.2008	104	16	-9	-4	-26	0,920
8.2008	98	17	-6	3	7	0,942
9.2008	132	18	34	40	37	1,347
10.2008	153	19	21	-13	-53	1,159
...

Zdroj: spracovanie vlastné pomocou MS Excel na základe[PACÁKOVÁ, 2003]

Obr. 2: Vývoj počtu prekladov z anglického jazyka do slovenského jazyka



Zdroj: spracovanie vlastné pomocou MS Excel

Z tabuľky č.3 a z obrázku č.2 vyplýva, že vývoj počtu prekladov z anglického do slovenského jazyka obsahuje trendovo–cyklickú zložku. Pre mechanické vyrovnanie časového radu sme používali metódu kĺzavého priemeru (Pacáková, 2003).

Časový rad sme upravili metódou kĺzavého priemeru, pričom dĺžku kĺzania sme stanovili na kvartálnej báze, čiže sme použili $k=4$. Metódou kĺzavých priemerov z vyrovnaného časového radu sme vylúčili cyklickú zložku časového radu. Časový rad kĺzavých priemerov predstavuje len trendovú zložku. Takto upravený časový rad sme ďalej analyzovali pomocou exponenciálneho trendu, za účelom prognózy pre nasledujúce 4 mesiace. Pri prognóze sme vychádzali z počtu prekladov evidovaných v posledných troch mesiacoch. Na základe analýzy predpokladáme, že počet objednávok bude ako uvádzame v tabuľke 4.

Tab. 4 Prognóza počtu objednávok (prekladov z anglického jazyka do slovenského)

Čas (mesiac a rok)	Počet predpokladaných objednávok
10.2012	69,32833707
11.2012	70,33332936
12.2012	71,33832165
01.2013	72,34331394

Zdroj: spracovanie vlastné pomocou IBM SPSS, MS Excel

Záver

Na základe zhlukovej analýzy databázy zákazníkov klientov nemenovanej jazykovej agentúry môžeme rozdeliť na tri segmenty. Posledný z nich (tretí segment) disponuje charakteristikami cieľovej skupiny. Pre jazykovú agentúru preto odporúčame orientovať sa na zákazníkov, ktorí pochádzajú z Českej republiky z mesta Brno a pracujú v oblasti obchodu a dopravy, alebo sa zaoberajú nehnuteľnosťami a prenájmom. Vzhľadom na rastúci spracovateľský priemysel by bolo rozumné zabezpečiť spoluprácu s prekladateľmi z oblasti techniky a chémie, aby vedeli uspokojiť potrebu firiem v tejto oblasti. Z χ^2 testu vyplýva, že zákazníci cielene volia prekladateľské editory na základe odboru prekladu. Z tohto dôvodu by mala firma zabezpečiť spoluprácu s takými dodávateľmi, ktorí majú nielen potrebné jazykové zručnosti a špecializáciu ale majú dispozíciu aj špeciálne prekladateľské editory, ktoré sú nepostrádateľnými pre vyhotovenie odborných prekladov. Ako príklad by sme uviedli technické preklady, pri ktorých zákazník s najväčšou pravdepodobnosťou objedná preklad v nasledujúcich editoroch: MS Word 47,3%, NEMO 23,3%, Trados TagEditor 6,9%.

Štatistické metódy pomôžu firmám jednak prognózovať predaj v nasledujúcich mesiacoch na základe minulých období, na druhej strane dataminingové techniky pomôžu pochopiť

Reprodukce lidského kapitálu – vzájemné vazby a souvislosti. 9. – 10. prosince 2013

správanie sa zákazníka. Zistenie zo zhlukovej analýzy pomáha firmám nájsť cieľových zákazníkov. Následne firma môže posilniť svoju marketingovú komunikáciu a svoj produktový rad, ktorý sa týka určitej cieľovej skupiny. Ak si spoločnosť uvedomí potrebu zákazníka a zároveň zabezpečí, aby potreba zákazníka bola uspokojená, môže očakávať, že posilní svoj vzťah so zákazníkmi. To v konečnom dôsledku vedie k silnejšiemu postaveniu firmy na trhu jazykových služieb. Ale musíme si byť vedomí toho, že ani štatistické a dataminingové techniky nevyriešia všetky problémy firiem. Faktory externého prostredia ako ekonomické, technické, sociálne, demografické, politické, právne a ekologické sú neodvratné a preto potrebuje podnik neustále sledovať svoje okolie, byť pripravený na zmenu trhu, požiadavky zákazníkov a zabezpečiť kvalitných dodávateľov.

Literatúra

- [1] ARMSTRONG, G. – KOTLER, P. 1990. Marketing. Bratislava: Slovenské pedagogické nakladateľstvo.
- [2] BERRY, J.A.M. – LINOFF, S. G. 2004. Data mining techniques: for marketing, sales, and customer relationship management. Indianapolis, Wiley Publishing, Inc.
- [3] KURDYOVÁ, E., 2013. Marketingové kampane a analýza dát. Bakalárska práca FM UK, Bratislava (školiťel: Bohdalová, M.)
- [4] PACÁKOVÁ, V. a kol. 2003. Štatistika pre ekonómov. Bratislava, Iura Edition.
- [5] RUBLÍKOVÁ, E., 2007. Analýza časových radov. Bratislava, Iura Edition.
- [6] ŘEZANKOVÁ, H. – HÚSEK, D. – SNÁŠEL, V. 2009. Shluková analýza dat. Praha, Professional Publishing.
- [7] VERCELLIS, C. 2009. Business Intelligence: Data Mining and Optimization for Decision Making. Chichester, Wiley Publishing, Inc.

Kontakt:

doc. RNDr. Mária Bohdalová, PhD.

Fakulta managementu UK v BA

Odbojárov 10

820 05 Bratislava

Maria.bohdalova@fm.uniba.sk

Bc. Eva Kurdyová, 4. ročník

Fakulta managementu UK v BA

Odbojárov 10

820 05 Bratislava

Eva.kurdyova@st.fm.uniba.sk