

Využitie technológie Big Data v praxi a na analýzu fraudov

The logo consists of a lowercase orange 'e' followed by the word 'FOCUS' in a bold, white, sans-serif font. A horizontal orange line is positioned below the text.

eFOCUS

Logo partnera

Využitie technológie Big Data v praxi a na analýzu fraudov

eFOCUS

Marek Zeman, Tatra banka, a.s.



° Čo je Big data – filozofický úvod



- ° Veľké komplexné dátové zdroje, navzájom ťažko spojitelné.
- ° Charakteristiky:
 - Objem
 - Rýchlosť
 - Rôznorodosť
 - Vierohodosť, variabilnosť, komplexnosť,.....
- ° Nutná podmienka ľudská „prirodzenosť“ pri analýze.
- ° Informácie môžu nie byť dáta z rozdielnych zdrojov.
 - Napr. správanie z proxy-logov prináša informácie o podobnom správaní jedincov (behaviour patterns).

° Big data čo to je?



- ° Koncepcia?
 - Zmena z priamočiarych ku nepriamočiarym vzťahom
 - Objem (hovoríme aspoň o terabytoch?)
 - Rýchlosť (naozaj potrebujete reálny čas?)
 - Rôznorodosť (aká je miera rôznorodosti?)
- ° Ako na to?, alebo ani Hadoop vám nepomôže.
- ° Pri vyhodnocovaní musí byť človek.

° Prečo je téma horúca práve teraz?



- ° Nové, neodkryté biznis možnosti.
- ° Väčšie zblíženie sa so zákazníkmi.
- ° Komplexnejší pohľad na správanie ľudí.

- ° Utilizácia zdrojov.
- ° Prioritizácia vývoja.

- ° Výhoda
 - Častočné pochopenie správania klienta.
 - Cielený marketing.

° Koncepty využitia?



- ° Kampanena Call centrum
 - Inbound
 - outbound
- ° Vyšperkované mobilnej aplikácie
- ° Dokonalý Internet banking
- ° Emailové kampane
- ° CRM
- ° Reklamácie

- ° Vnutro firmy
 - Zdokonaľenie interných aplikácií
 - Fraud solution

Vrátíme sa ku konceptom neskôr.

° Čo je vlastne na koncepte nové?



Nestačí BI?

Nestačí!

- ° Pridanie dát, ktoré sa bežne nespracovávajú.
- ° Vytvorenie kompletného pohľadu na klienta.
- ° Pridanie externých zdrojov.
- ° Rovnaké typy informácií (firmy ako aj fyzických osobách)
- ° Inštalovanie a spracovávanie komplexného pohľadu na klienta vyšetrovacím SW.

° Popísanie zdroja dát



- ° Štandardné
 - Bankové (CBSs, Clientské, IB, DWH, ODS, ...)
 - Používateľské (LN, IPCC, proxy)

- ° Neštandardné
 - Systémové
 - Sieťové (FW, WAF, IDS, AF, ...)
 - Bezpečnostné
 - Externé zdroje

° Vieme kde sú dáta a?



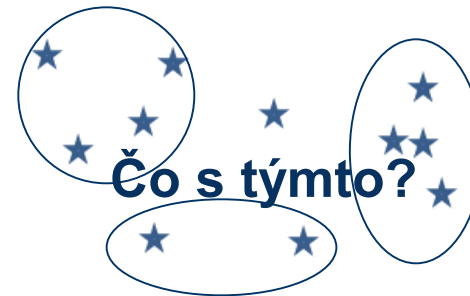
- ° Centrálné úložiská (áno aj nie – je toho veľa, načo?)
- ° Rôzne typy (štruktúry, formy, zdroje)
- ° Rôzne prístupy (BI, Reporty, Text mining,....)



° Dáta máme, rozumie im niekto?



- ° Štandardné BI prístupy
- ° Štandardné Reporty
- ° Neštandardné prístupy
 - Klasifikácia či klastering
 - Overovanie klastrov
 - Korelácia verzus kauzalita
 - Štatistika, štatistika, štatistika
 - Dobré nápady = ľudia



° Ako by to je a ako by to mohlo byť



Tatra banka Security web

Telefonické hovory



[Domov](#) [Novinky](#) [Know-How](#) [Služby a evidencie](#) [Projekty](#) [OBIS a NT](#) [Rôzne](#) **[CeMoL](#)**

CeMoL » Telefonické hovory

Tel. hovory

Najčastejšie volania

Kontakty

Service Desk: 3333
Produktový Support: 2000

RSS Novinky

[RSS - novinky](#)

Užívateľ: [REDACTED]

Telefonické hovory v čase od 1.9.2014 do 30.9.2014

Volajúci	Meno	Volaný	Meno	Čas	Dĺžka hovoru
3955	[REDACTED]	0903789657	[REDACTED]	2014-09-04 14:13:49.0	00:00:17
421912193955	[REDACTED]	421903789657	[REDACTED]	2014-09-04 16:13:00.0	00:00:18
2520	[REDACTED]	0903789657	[REDACTED]	2014-09-05 06:55:40.0	00:03:38
421912192520	[REDACTED]	421903789657	[REDACTED]	2014-09-05 08:55:00.0	00:03:39
1682	[REDACTED]	0903789657	[REDACTED]	2014-09-09 11:59:17.0	00:04:02
421912191682	[REDACTED]	421903789657	[REDACTED]	2014-09-09 13:59:00.0	00:04:03
0917778221	[REDACTED]	00421903789657	[REDACTED]	2014-09-16 12:49:00.0	00:02:11
0905399084	[REDACTED]	00421903789657	[REDACTED]	2014-09-18 11:54:00.0	1
0905399084	[REDACTED]	00421903789657	[REDACTED]	2014-09-18 11:54:00.0	1
3127	[REDACTED]	0903789657	[REDACTED]	2014-09-18 12:00:45.0	00:00:11
1992	[REDACTED]	0903789657	[REDACTED]	2014-09-18 12:08:31.0	00:00:49
421912193127	[REDACTED]	421903789657	[REDACTED]	2014-09-18 14:00:00.0	00:00:11

° Konceptia Call centrum.



- ° Analýza klientov
- ° Pripravené ponuky
- ° Inbound hovory
- ° Outbound hovory

- ° Dá sa ešte zlepšiť?



° Konceptia CRM a reklamácie



- ° Spracovanie reklamácií:
 - Identifikovanie témy reklamácie
 - Identifikovanie názoru na produkt
 - Meranie polarity názorov
 - Meranie miery názoru
- ° CRM
 - Téma rozhovoru
 - Plánovaný rozvoj
 - Ďalšia interakcia

Text mining.



° Konceptcia: IB a aplikácie



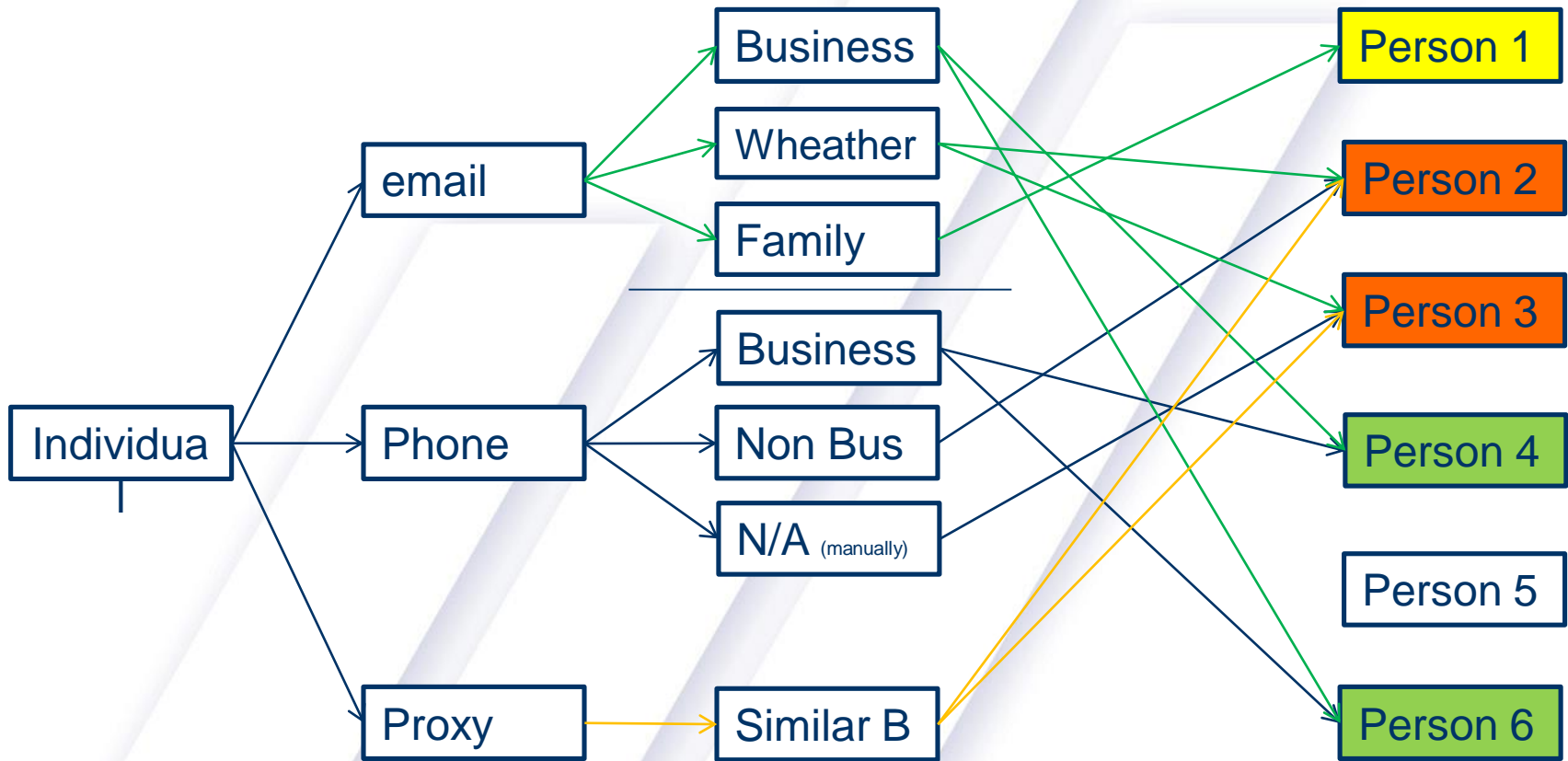
° Filozoficky....

- Dá sa vyberať a dohodnúť špecifická vzorka na testovanie
- Porovnať si, či výber korešponduje so správaním



Students by Source	Step 1 Enrolled	Step 2 Enrolled - online Student	Step 3 Enrolled	Step 4 Enrolled - Study + Tests	Step 5 Enrolled - Study + Tests	Step 6 Admitted	Step 7 Enrolled - online Student	Step 8 Admitted - online Student Tests Passed
group	24.3K - students	20.8K - 85% of 24.3K	17.9K - 71% of 21.2K	1.8K - 10% of 21.2K	1.8K - 10% of 21.2K	700 - 3% of 21.2K	700 - 3% of 21.2K	600 - 3% of 21.2K
International	11.8K - students	11.3K - 95% of 11.8K	9.2K - 78% of 11.8K	1.9K - 18% of 11.8K	1.8K - 14% of 11.8K	800 - 6% of 11.8K	800 - 6% of 11.8K	600 - 5% of 11.8K
Local	8.27K - students	8.27K - 100% of 8.27K	8.8K - 10% of 8.27K	128 - 2% of 8.27K	138 - 2% of 8.27K	40 - 0% of 8.27K	40 - 0% of 8.27K	40 - 0% of 8.27K
Non-UK	354 - students	354 - 100% of 354	84 - 24% of 354	0 - 0% of 354	0 - 0% of 354	0 - 0% of 354	0 - 0% of 354	0 - 0% of 354
UK	100 - students	100 - 100% of 100	31 - 31% of 100	0 - 0% of 100	0 - 0% of 100	0 - 0% of 100	0 - 0% of 100	0 - 0% of 100
...	1.8K - students	1.8K - 100% of 1.8K	377 - 21% of 1.8K	127 - 7% of 1.8K	117 - 7% of 1.8K	40 - 2% of 1.8K	40 - 2% of 1.8K	31 - 1% of 1.8K
Total	40.3K - students	35.8K - 89% of 40.3K	24.2K - 59% of 40.3K	3.8K - 9% of 40.3K	3.6K - 9% of 40.3K	1.8K - 4% of 40.3K	1.8K - 4% of 40.3K	1.3K - 3% of 40.3K

° Fraud solution

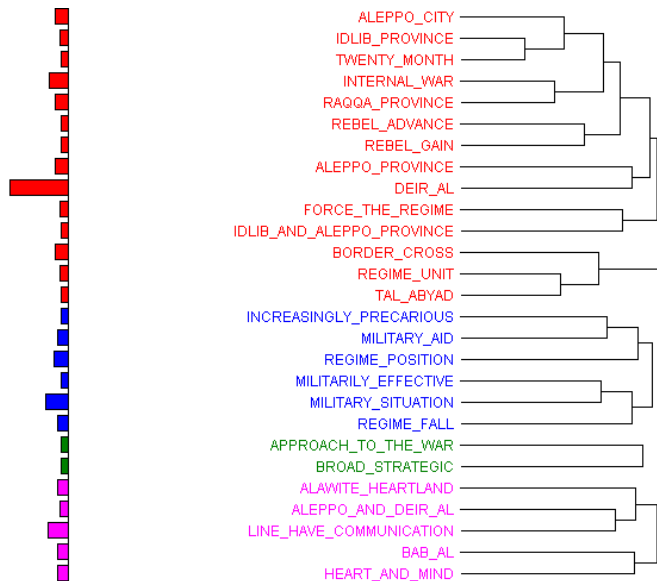




?

Clustering

Data – Syria 2011/2013 , source: www.washingtoninstitute.org



EXAMPLE

TARGET	KEYWORD	CO-OCCURS	DO NOT	IS ABSENT	Jaccard	STRENGTH
AIR_MISSILE	ANTI-AIRCRAFT_GUN	4	1	4	0,444
AIR_MISSILE	WEAPON_SYSTEM	4	2	4	0,400
AIR_MISSILE	ANTI-AIRCRAFT_WEAPON	3	1	5	0,333
AIR_MISSILE	FIELD_ARTILLERY	4	4	4	0,333
AIR_MISSILE	SYRIAN_AIR_FORCE	3	1	5	0,333
AIR_MISSILE	CIVILIAN_CASUALTY	4	5	4	0,308
AIR_MISSILE	CW_CAPABILITY	3	2	5	0,300
AIR_MISSILE	CHEMICAL_ATTACK	4	6	4	0,286
AIR_MISSILE	RECOILLESS_RIFLE	2	0	6	0,250
AIR_MISSILE	AIR_DEFENSE	3	4	5	0,250
AIR_MISSILE	INTERNAL_WAR	3	5	5	0,231
AIR_MISSILE	REBEL_ADVANCE	2	1	6	0,222
AIR_MISSILE	WING_AIRCRAFT	2	1	6	0,222
AIR_MISSILE	SURFACE_MISSILE	3	6	5	0,214
AIR_MISSILE	COMMAND_AND_CONTROL	4	11	4	0,211

Ďakujeme za pozornost

eFOCUS