

Názov univerzity: **Technická univerzita v Košiciach**
Názov fakulty: **Fakulta elektrotechniky a informatiky**
Názov katedry: **Katedra kybernetiky a umelej inteligencie**
Akademický rok **2014/2015**
Stupeň vysokoškolského štúdia, ročník: **III. stupeň, 1. ročník**
Lektor z vysokej školy: **prof. Ing. Iveta Zolotová, CSc.**

ZÁVEREČNÁ PRÁCA
ZO VZDELÁVACIEHO POBYTU V ZAHRANIČÍ
realizovaného s podporou národného projektu „Vysoké školy ako motory
rozvoja vedomostnej spoločnosti“, ITMS kód projektu 26110230120

Meno a priezvisko študenta: **Martin Miškuf**
Miesto realizácie vzdelávacieho pobytu v zahraničí: **Socialbakers a. s., Praha, ČR**
Kontaktná osoba: **bc. Michaela Egedová**
Obdobie realizácie vzdelávania: **2.8.2015 – 1.9.2015**

Miesto realizácie vzdelávania v podmienkach podnikovej praxe na Slovensku:
U.S. Steel Košice, s.r.o.

Lektor z podniku: **Ing. Michal Smoleň**

Obsah

1. Úvod	3
2. Stanovené ciele vzdelávacieho pobytu v zahraničí	3
3. Priebeh praxe (metodika plnenia cieľov)	3
3.1. Detekcia fake followerov na Tweeteri pomocou Deep Learningu	4
3.2. Analýza českého a slovenského internetu v nástroji Social Insider	6
4. Sumarizácia (naplnenie cieľov).....	7
5. Odporúčania	7
6. Záver	7
7. Prílohy.....	7

1. Úvod

V auguste som sa zúčastnil stáže vo firme Socialbakers a.s., v Prahe v Českej republike, ktorá sa zaoberá analýzou dát zo sociálnych sietí. Táto firma sa behom posledných rokov rozrástla a v súčasnosti (2015) má okolo 350 zamestnancov po celom svete. Jej zakladateľ Jan Rezac bol magazínom Fortune zaradený do celosvetového rebríčka 30 ľudí pod 30 v oblasti sociálnych sietí a marketingu. Stáž prebiehala v období od 2.8.2015 (nedeľa) do 1.9.2015 (utorok).

2. Stanovené ciele vzdelávacieho pobytu v zahraničí

V spoločnosti Socialbakers som bol pridelený na oddelenie R&D (research and development), kde sa po konzultáciách v harmonograme vyšpecifikovali nasledovné oblasti:

- Princípy analýzy sociálnych sietí s pomocou interných nástrojov Socialbakers a programovacieho jazyka R
- Monitoring a z toho vyplývajúce analýzy českého a slovenského internetu (v nástroji Social Insider)
- Analýzy a zpracování strategií značek na příslušných sociálních médiích (Facebook, Twitter, YouTube)

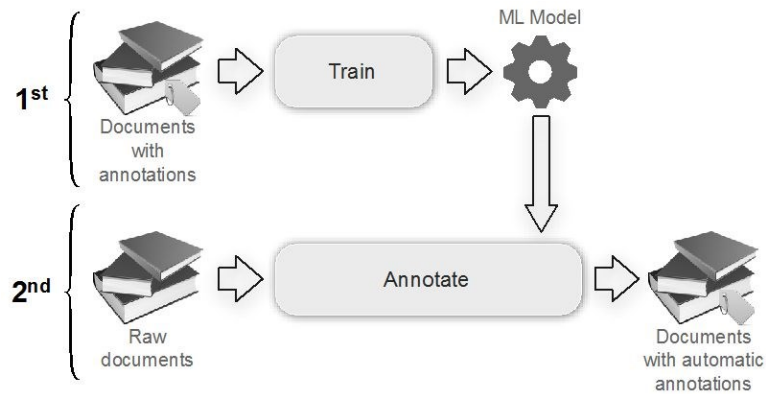
3. Priebeh praxe (metodika plnenia cieľov)

Po príchode na miesto výkonu praxe (3.8.2015 pondelok), mi Michaela Egedova z HR oddelenia urobila úvodnú inštrukciu, ukázala mi priestory spoločnosti, dala mi podpísať zmluvu o mlčanlivosti – preto nebudem v tejto správe špecifikovať technické detaily. V spoločnosti Socialbakers ma prideliť na oddelenie R&D (research and development). Toto oddelenie sa zaoberá vylepšovaním produktov a vývojom nových produktov. Tento vývoj prebieha buď na požiadavku zákazníka alebo sledovaním požiadaviek trhu a ponuky konkurencie. Vedúcim tohto oddelenia je pán Josef Šlerka (na fotke v fotodokumentácii) a okrem neho na tomto oddelení pracujú ďalší 4 ľudia.

V prvých dňoch stáže mi dali za úlohu oboznámiť sa s programovacím jazykom R, ktorý sa využíva na analýzu dát. Tento jazyk je obdobou jazyka Matlab, ale je zadarmo. Ponúka možnosť doinštalovania rôznych balíčkov ktoré obsahujú funkcie napríklad na prácu so sociálnymi sieťami ako Twitter, Facebook, Instagram ... a ďalšie balíčky špecializujúce sa na prácu s dátami ako napríklad balík H2O od spoločnosti Coursera, ktorý ponúka funkcie z oblasti neurónových sietí / Deep learningu.

Po oboznámení s jazykom R a vývojovým prostredím R štúdia mi vedúci oddelenia vyšpecifikoval úlohu detegovania falošných followerov za pomoci strojového učenia, presnejšie Deep learningu na sociálnej sieti Twitter.

3.1. Detekcia fake followerov na Tweeteri pomocou Deep Learningu

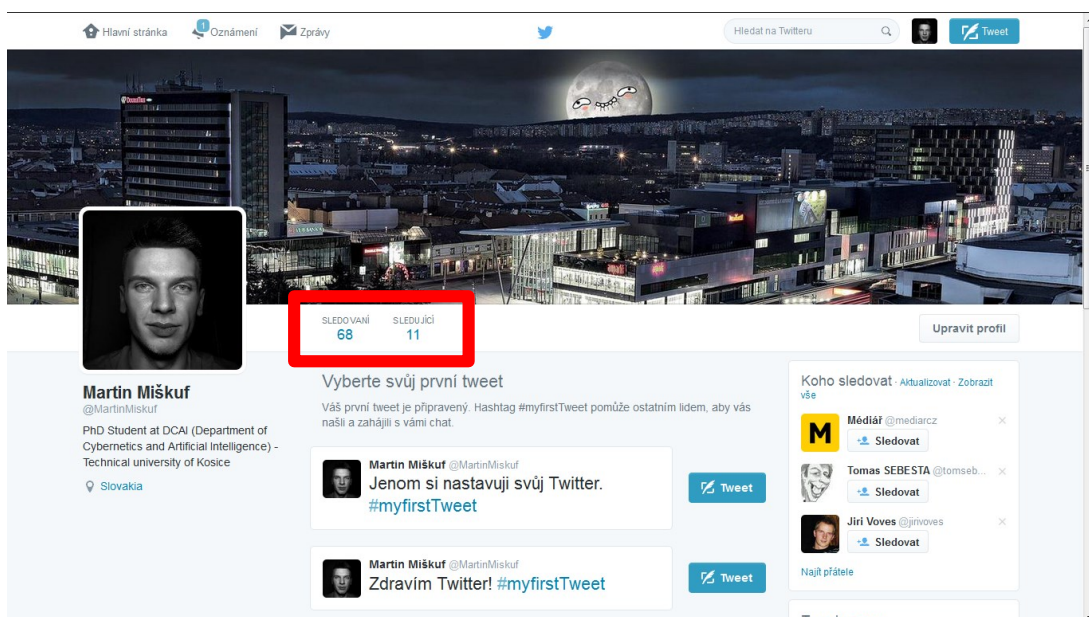


Princíp nekontrolovaného strojového učenia

Pri strojovom učení pomocou neurónových sietí som musel prejsť týmito krokmi:

- Potrebujeme dáta ktoré budú zoznamené podľa nejakej vlastnosti
- Potrebujeme neurónový model ktorý sa môže s týmito dátami natréňovať a zistíme jeho presnosť
 - ak je dobrá použijeme ho
 - ak je presnosť zlá opravíme parametre modelu
- Dobrý model môžeme použiť na predpovede, respektíve rozdelenie dát

Sociálna sieť Twitter slúži na to aby používatelia mohli zdieľať / tweetovať ich myšlienky. Ak sa vám páčia tweety a ostatná činnosť iného používateľa môžete ho nasledovať / stanete sa followerom. To znamená že budete vidieť všetky jeho príspevky na vašej nástenke.



Ukážka profilu na sociálnej sieti Twitter so zvýraznením počtu sledovaných a sledujúcich

Vo všeobecnosti platí čím viac followerov ma používateľ tým je vplyvnejší. Sú tu však internetové spoločnosti ktoré ponúkajú možnosť kúpenia si followerov a takýto followeri sú označovaní ako fake (falošný) followeri.

Get More Followers = Get More Visitors + Traffic + Customers

Free Service: Get Free Twitter Followers	Premium Service: Get Thousands Followers
Get Follower Free Every Day ✓	Get Follower Free Every Day ✓
Add 1000 followers immediately. (First Day) ✗	Add 1000 followers immediately. (First Day) ✓
Autopilot Get Follower (RealTime) ✗	Autopilot Get Follower (RealTime) ✓
Guarantee UnFollow Back ✗	Guarantee UnFollow Back ✓
Report Counter Onto Follower ✗	Report Counter Onto Follower ✓
Become Affservice ✓	Become Affservice ✓

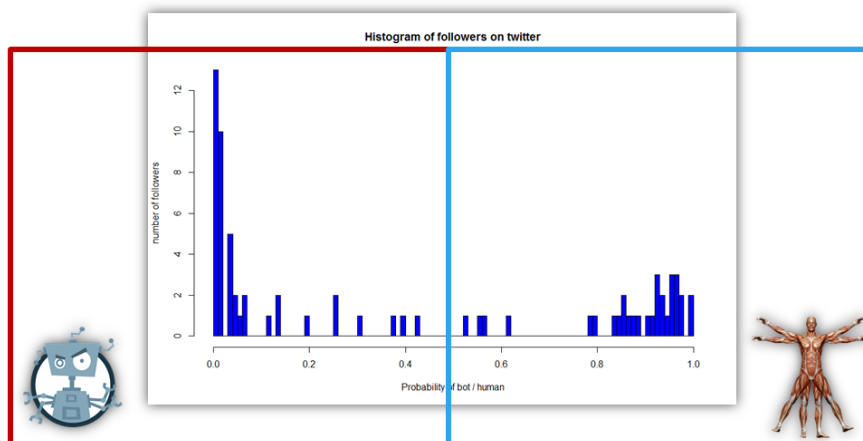
Start Now! **Buy Now!**

Takto sa dajú nakúpiť falošný followeri

Ak sa takýto followeri vyskytnú u nejakého používateľa vo veľkom množstve, pomere ku skutočným ľudským followerom je to podozrivé. Týmto spôsobom sa da skresľovať vplyv používateľov na Tweeteri, preto je to zaujímavá úloha pre spoločnosť zaoberajúcu sa dátovou analytickou zo sociálnych sietí.

Mojou úlohou bolo naučiť sa robiť so spomínanými balíčkami pre jazyk R (twitterR, H2O). Pomocou jazyka R som dokázal stiahnuť followerov pomocou Rest API ktoré Twitter ponuka. Tých bolo následne potrebné odabovať či sú to boti (používateľa vytvorený umelo) alebo ľudia (účty skutočných ľudí ktorí ich používajú na tweetovanie a followujú používateľov ktorí sa im páčia), to znamená bolo potrebné vytvoriť dataset na ktorom sa neurónová sieť mohla učiť.

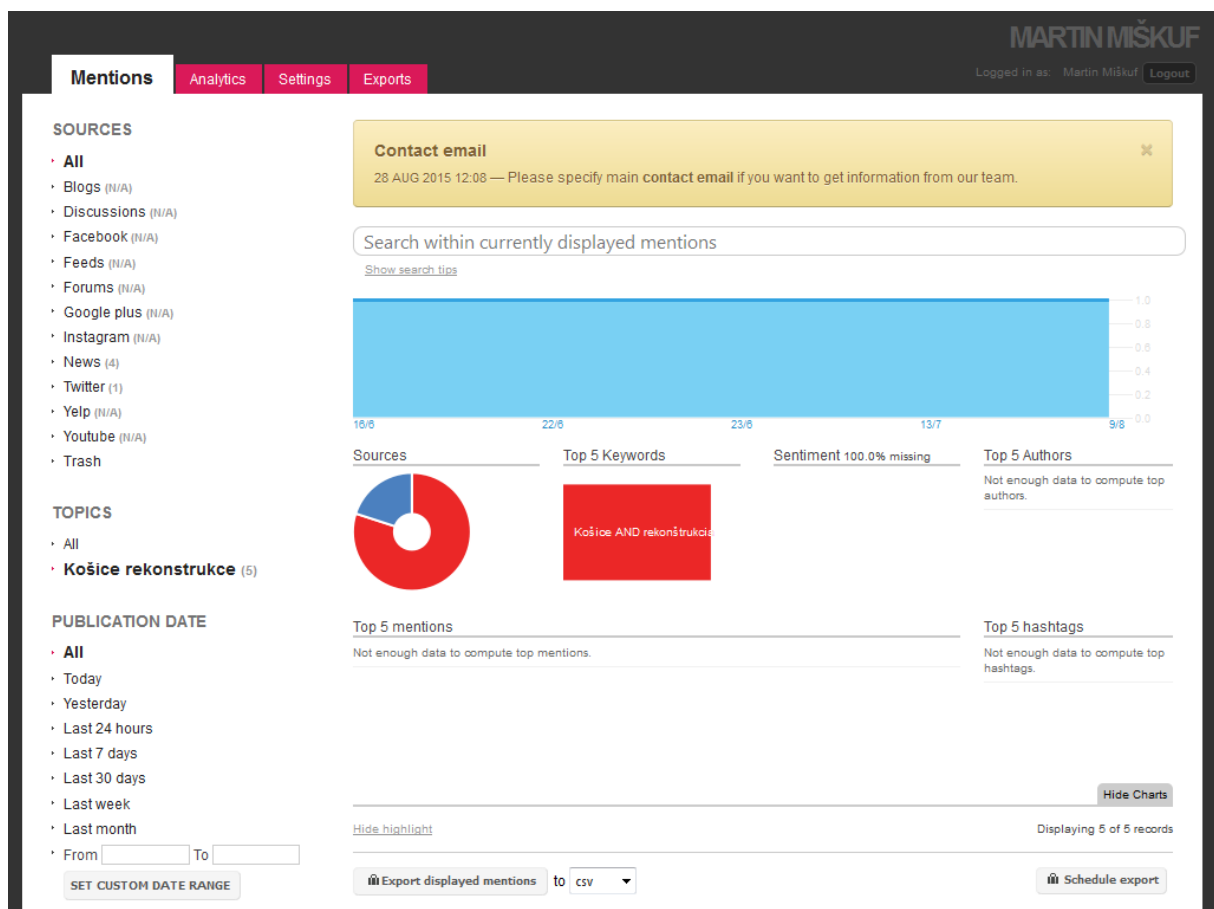
Následne som mohol nakonfigurovať a natréňovať neurónovú sieť ktorá po natréňovaní dosahovala 84% úspešnosť v detekcii botov. Výstupom bol model do ktorého sa mohli nahráť dáta a pre tieto dáta, pre každého používateľa bola vypočítaná pravdepodobnosť toho že je to človek.



Výstup z môjho riešenia je histogram kde vidno rozdelenie používateľov podal pravdepodobnosti

3.2. Analýza českého a slovenského internetu v nástroji Social Insider

Okrem práce s jazykom R a neurónovými sieťami som sa na stáži oboznámil s nástrojom Social Insider, ktorý slúži na analýzy dát zo sociálnych sietí so zameraním na československý trh. Tento nástroj monitoruje celý Česko-Slovenský sociálny priestor – Facebook, Twitter, zápravy a mnohé iné informačné kanály a tieto dáta ukladá. Je v podstate business intelligence riešenie, ktoré ponúka užívateľom nástroje na vytváranie reportov. Tento produkt je spolplatnený podľa oficiálneho cenníka spoločnosti Socialbakers.



Screenshot z obrazovky kde som skúšal nájsť všetky zmienky ktoré obsahujú kľúčové slova Košice, rekonštrukcia, trať, električka

Reporty z tohto nástroja sa môžu exportovať do rôznych formátov a ďalej spracovávať v nástrojoch napríklad od spoločnosti Microsoft – Office – Excel, Word ... Takýto report z trial verzie ktorý obsahuje zmienky za posledné 3 dni si môžete pozrieť aj v prílohe.

4. Sumarizácia (naplnenie cieľov)

Počas stáže som si prešiel týmito krokmi:

- Naučil som sa programovať v jazyku R a pracovať s nástrojom R studio
- Naučil som sa pracovať s neurónovými sieťami a využívať ich pri tvorbe modelov pomocou ktorých sa môžu dáta analyzovať
- Vytvoril som aplikáciu ktorá slúži na detekciu falošných followerov na sociálnej sieti
- Na konci stáže som toto svoje riešenie od prednášal ľuďom z obchodného oddelenia ako výstup R&D oddelenia
- Ukázali mi ich spôsoby analýzy dát zo sociálnych sieti – Social Insider

Na základe týchto bodov môžem vyhlásiť, že stanovené ciele stáže z harmonogramu boli naplnené.

5. Odporúčania

Vďaka projektu som mal možnosť robiť v medzinárodnej spoločnosti ktorá už má vybudované svoje meno vo svete. Mal som možnosť porovnať svoje vedomosti s požiadavkami trhu a absolvovať rôzne školenia a získať ďalšie odborné vedomosti a skúsenosti.

6. Záver

Kvôli mojim pracovným povinnostiam moja stáž trvala iba jeden mesiac. Tento mesiac mi dal plno nových skúsenosti po odbornej stránke, ktoré využijem vo svojom ďalšom štúdiu. Okrem toho som však mal možnosť spoznať prácu v medzinárodnom kolektíve ľudí v pokrokovej dynamicky sa rozvíjajúcej firme.

7. Prílohy

Ukážkový report z nástroja Social Insider:

`export_martinmiskuf_2015-08-28-12-21_kosice-rekonstrukce.docx`

Fotodokumentácia praxe



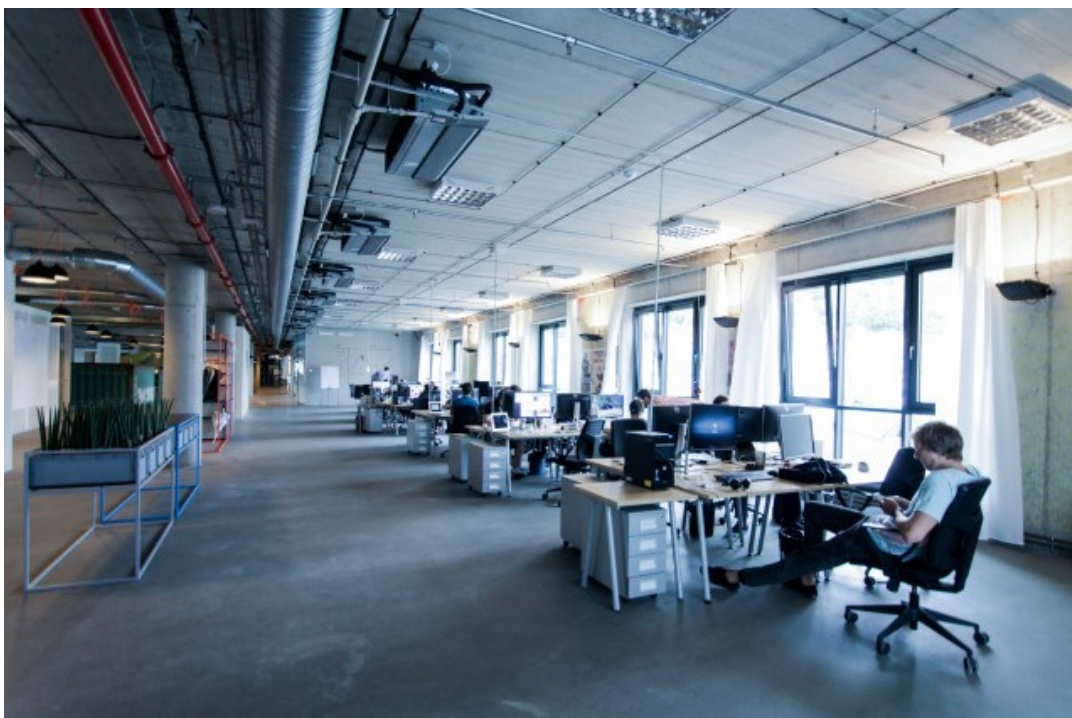
Môže sa zdať, že ste na stavenisku, pritom ide o premyslený koncept. Táto recepcia sa nachádza hneď oproti hlavnému vchodu v treťom poschodí.



V marketingovom oddelení je dominantou command centrum, kde sa v priamom prenose sledujú aktuálne metriky.



Zasadačky sú označené logami sociálnych sietí, čo je pre Socialbakers typické. Nájdete tu tak LinkedIn, Twitter alebo Spotify.



V každom poschodí sú väčšie priestory s podlhovastými stolmi. "Chceli sme vytvoriť open space, v ktorom má každý svoje miesto"



Orientáciu po dvoch poschodiach kancelárií uľahčujú na zemi nakreslené šípky.



Pohľad na jeden z open space - toto konkrétne patrí produktovému oddelenia.



**Netypickým prvkom sú kuchynky umiestnené v kontajneroch.
Nachádzajú sa v každom poschodí.**



**Za recepciou si môžete posediť v miestnej kaviarni, ktorá je spojená s kuchynkou
zabudovanou opäť v kontajneri.**

Fotodokumentácia označenia priestorov realizácie vzdelávacieho pobytu



Fotka pri práci s dátami v R štúdiu s neurónovými sieťami



Moja fotka s vedúcim Research and Development oddelenia – Josefom Šlerkom
V pozadí vidno označenie priestoru
(open space kancelárie)



Moja fotka s názvom spoločnosti – Socialbakers (v Prahe v Karlíne)



**Na konci stáže sme zorganizovali prednášku, kde som prezentoval na čom som na stáži pracoval
Jednalo sa o použitie Deep learningu na detekciu fakeovych followerov na Twitteri**



Fotka z označenia vchodových dverí

Súhlasím so zverejnením správy na komunikačnom portáli národného projektu pre potreby monitorovania a popularizácie projektu, prípadne na príslušných médiách vysokej školy, kde študujem.

Podpis študenta:

Dátum odovzdania:

Kancelária národného projektu „Vysoké školy ako motory rozvoja vedomostnej spoločnosti“ vyhlasuje, že za obsah produktu zodpovedá výlučne autor a Európska komisia ani kancelária národného projektu nenesú zodpovednosť za akékoľvek použitie informácií obsiahnutých v danej záverečnej správe.