

Analýza nestrukturovaných dat z bankovních stránek na sociální síti Facebook

Unstructured Data Analysis from Facebook Banking Sites

Lucie Šperková¹

¹ Katedra informačních technologií, Fakulta informatiky a statistiky

Vysoká škola ekonomická v Praze

nám. W. Churchilla 4, 130 67 Praha 3

lucie.sperkova@vse.cz

Abstrakt: Sociální sítě jako novodobý kanál pro sdílení online Word of Mouth představuje pro bankovní společnosti příležitost, jak analyzovat komunikaci a názory svých zákazníků, které zde veřejně sdílejí. Článek ukazuje proces a výsledky pokročilé kontextové analýzy těchto nestrukturovaných dat vybraných veřejných bankovních stránek na sociální síti Facebook. Analýza zákaznických názorů, problémů a přání může hrát významnou roli pro pozici těchto společností na trhu a jejich orientaci na zákazníky. Využity byly metody shlukování a sentiment analýz, výsledky odrážejí stav chování zákazníků a bank na této síti. Článek zdůrazňuje důležitost problematiky Word of Mouth a jeho analýz informačními technologiemi. Zároveň diskutuje, jak analýza kontextu může přenést marketing na další úroveň identifikace zákaznickova chování a jeho vztahu ke společnosti.

Klíčová slova: Nestrukturovaná data, Facebook, shlukování, sentiment analýza, Voice of Customer, Word of Mouth, banka.

Abstract: Social networks as a modern channel for sharing online Word of Mouth is an opportunity for banking companies to analyse the communications and opinions of their customers, which they are sharing publicly. The article shows the process and results of these advanced contextual analyses of unstructured data of selected public banking pages on social networking site Facebook. The analysis of customer opinions, problems and desires may play an important role for the position of these companies in the market and their orientation to customers. Clustering methods and sentiment analysis were used, results reflect the state of the behaviour of customers and banks in this network. Paper highlights the importance of the issue of Word of Mouth and its analyses by information technology. At the same time discusses how context analysis can migrate marketing to the next level of identification of customer behaviour and his relationship to the company.

Keywords: Unstructured data, Facebook, Clustering, Sentiment analysis, Voice of Customer, Word of Mouth, bank.

1 Úvod

S neustále vyšším používáním sociálních sítí mezi populací je generováno také čím dál větší množství nestrukturovaných dat. Ta mohou obsahovat vysoký potenciál pro získávání informací a marketingový výzkum. Zůstat konkurenceschopný v těchto podmínkách trhu znamená být více orientovaný na zákazníky a vést efektivní sociální marketingové kampaně. Kritickým faktorem tohoto úsilí je primárně analýza Voice of Customer (zkratka VoC), takzvaného hlasu zákazníka, jindy nazývaného jako Word of Mouth (zkratka WoM), v překladu slovo z úst. Přestože se tyto pojmy v literatuře zaměňují a je jimi chápáno to samé, lze v jejich definicích nalézt určité rozdíly.

Hu et al., (2006) chápou WoM je veškerou neformální komunikaci směřovanou dalším konzumentům o vlastnictví, používání nebo charakteristikách daného produktu či služby nebo jejich prodejců, tedy společnosti a značky. Helm a Schlei (1998) definovali WoM marketing jako monitorování ústní komunikace, pozitivní i negativní, mezi zákazníky, dodavateli, nezávislými experty o značce, produktech, službách a dalších věcech týkajících se vybrané firmy. Tato druhá definice se ovšem více blíží definici Voice of Customer, kterému se věnují (Griffin a Hauser, 1993). Ti tvrdí, že cílem miningu VoC (dnes bychom řekli opinion miningu) je pochopení zákaznických potřeb a jejich transformace do klíčových funkčních požadavků. Neříkají tedy nic o doporučování, hlavním záměrem většiny výzkumů WoM na internetu. Helm a Schlei (1998) rozdělují WoM na spontánní, které odpovídá definici VoC, organické, jímž je míněno ono doporučování a dále WoM amplifikované, jež je výsledkem marketingových aktivit cíleně podporujících šeptandu mezi lidmi. Zjednodušeně tedy lze říci, že pokud analyzujeme veškerý obsah, který uživatel napsal či řekl, analyzujeme tak jeho „hlas“, tedy VoC. Jeho vliv na další účastníky tohoto „hovoru“ lze pak řadit mezi WoM.

Internetová verze WoM, v literatuře nazývaná také jako digitální (Hu et al., 2006), elektronické WoM (Choi a Scott, 2012) nebo online, se tak stává majoritním informačním zdrojem pro zákazníky. Jakýkoliv konzument na světě se může připojit na internet a přečíst si názory jiných. Tato komunikace má velmi významné důsledky pro širokou škálu manažerských činností, jako je například

- budování značky,
- získávání a udržení zákazníků,
- vývoj produktu,
- zajištění kvality.

WoM doporučení skrz internet a sítě jsou jednoduchá, zdarma, volně dostupná a relativně efektivní data v šumu informací, jež nás obklopují. Lidé mají tendence věřit názorům ostatních více než letákům ve schránce, což má také důležitý dopad na jejich rozhodovací procesy. Ty jsou WoM ovlivňovány nejvíce, když zákazníci chtějí koupit produkt poprvé, je relativně drahý, nebo když zcela ne úplně rozumí produktu nebo službě, kterou aktuálně spotřebovávají (Bughin et al., 2010). Poslední případ je vidět zejména v oblasti bankovníctví. V takovém případě mají tendence hledat si více informací a srovnávat mezi několika službami, protože podmínky nákupu v tomto případě nejsou pro spotřebitele jasné a pochopitelné. Sociální síť, poskytující spotřebitelům vhodnou platformu pro sdílení jejich názorů, tak představuje posun v komunikaci tzv. „jeden na jednoho“ na šíření informace z jednoho uzlu na několik, viz výzkum Han a Niu (2012). Stejně tak umožňuje organizacím kromě propagace také naslouchat požadavkům a přáním svých uživatelů a tak zlepšovat své služby a produkty.

Obecně využívanou praktikou marketérů je manuální identifikace klíčových zákaznických potřeb z masových dotazníků, výzkumů a rozhovorů. Tyto nahodilé, těžko škálovatelné a k chybám náchylné procesy jsou časově a zdrojově velice náročné. Celkově společnosti utrácí za marketing značné části svých rozpočtů, v posledních letech většinou na komplikovaně koncipované kampaně, které potom manuálně vyhodnocují. (Bughin et al., 2010) Na zákaznická data se také soustředí datově řízené analytické moduly CRM napojené na front-end aplikace. Většina těchto metod je založena na standardizovaných strukturovaných datech a relačních schématech, kdy datová pozice determinuje chování algoritmu, především na základě klasifikací nebo asociačních pravidel typu if-then. I tyto procesy vyžadují velké úsilí a stojí za nimi aktivity třídění, čištění a transformace nestrukturovaných dat do matic, proto nejsou pro kontinuální exponenciální růst nestrukturovaných dat vhodné. Relační modely nejsou schopny odpovídat na otázky vyžadující nestrukturovaná data, tedy dodávat indiskrétní odpovědi na otázky obsahu a sentimentu daného textu. Data na webu jsou dynamická již ze své podstaty, jak se mění díky neustálé aktualizaci a doplňování nových informací ze strany uživatelů. Neschopnost tato data analyzovat uživatelsky přívětivou cestou představuje nejen technologickou výzvu, ale i změnu myšlení pro marketéry, kteří si význam těchto analýz musí uvědomit a především porozumět, jak generovat koordinované, konzistentní, relevantní odpovědi ve správný čas těm správným lidem. Jedině tak lze ovlivnit zákaznicko rozhodování, doporučení, spokojenost a loajalitu.

1.1 Sociální marketing

Pravděpodobně nejpoužívanější metodou v digitálním sociálním marketingu je monitoring sociálních sítí. Obecně však výpočty návratnosti investic nelze uspokojit jen měřením počtu typu imprese či počtu fanoušků na síti. Pro hlubší zkoumání zákaznickovy loajality, chování a spokojenosti by data měla být čtena v širším kontextu, analyzován jejich obsah, sentiment a vztahy mezi daty a jejich kontextem. Dle Heudeckera (2013) nejsou sociální analýzy vyspělou technologií, přesto mají dobře identifikovány případy užití, které vysvětlují jejich hodnotu pro marketingový výzkum. Han a Niu (2012) provedli výzkum propagace WoM na sociálních sítích, ale cíl byl ukázat vliv různých uzlů v síti, nikoliv dolování kontextu. Navzdory popularitě sociální sítě Facebook, neexistuje moc prací řešící sentiment v této síti, např. (Birmingham et al., 2009). Více analyzovanými sítěmi jsou sítě mikroblovací (Go et al., 2009; Pak a Paroubek, 2010; Barbosa a Feng, 2010). Veeck a Hoger (2014) provedli studii s určením kroků pro projekt marketingového výzkumu pro bakalářské stadium analyzující sociální média.

Někteří dodavatelé analytiky již přidali do svých řešení nástroje pro sociální analýzy jako text miningové nástroje. Také open source komunity hledají jak využít analýzy obsahu pro lepší porozumění veřejného mínění a jeho sentimentu. Existují dodavatelé nabízející SaaS API pro analýzy extraktů nestrukturovaných dat v cloudu, které dohromady s velkým objemem informací dostupných na internet akcelerují vývoj a adopci na tomto poli. Pro konkurenceschopnost dodavatelů jednoduchého monitoringu je nutné se posunout k hlubším analýzám a speciálním aplikacím. (Heudecker, 2013)

1.2 Stav bankovníctví v ČR

Bankovníctví v České republice zažilo v letech 2007-2011 růst, co se týče nově příchozích hráčů na trh (ČNB, 2014) a reorganizaci trhu, kdy tyto nové nízkonákladové instituce (konkrétně mBank S. A., organizační složka, Fio banka a.s., Air Bank a.s., ZUNO BANK AG, organizační složka a Equa bank a.s.) přebírají klienty již zavedených bank, především velké čtyřky ČSOB, Komerční banky, České spořitelny a UniCredit Bank. Nedisponují sice

hustou pobočkovou sítí, jsou považovány spíše za banky internetové, ale jako vůbec první na českém trhu lákají prospekty na nízké či nulové poplatky, úrokové sazby a rychlé jednoduché transparentnější služby ve formě internetových aplikací. Za několik let existence se jim podařilo převést více než 1,4 milionů klientů (Nacher, 2014) a velké bankovní domy na to musí reagovat, pokud nechtějí nadále ztrácet své klienty.

Nové technologie přinesly do bankovního sektoru v posledních letech mnoho inovací. Klesá tak konkurenční výhoda velkých bank s hustou pobočkovou sítí, jejichž význam již není tak vysoký. Vícekanálová distribuce, digitalizace (Půlpánová, 2007), cloud banking, big data, komoditizace a sociální sítě jsou trendy udávající směr, kterým by se banky měly co nejdříve vydat. Právě sociální sítě mohou dle Morrissey (2012) bankám výrazně pomoci v posilování důvěryhodnosti a transparentnosti. Jejich prostřednictvím lze snadno šířit reklamu v podobě zákaznických doporučení.

Budování klientské loajality je v současných podmínkách, kdy je možno snadno a často měnit banku, pro tyto instituce ještě důležitější než dříve. Analýza Voice of Customer může být proto kritická pro udržení konkurenceschopnosti na finančním trhu. Potenciál skrývá identifikace klientských potřeb a zájmů, trendů v bankovníctví, oblastí, kde je nutné zlepšování služeb, kde a jakým zákazníkům pomáhat s jejich problémy a zejména jak s nimi komunikovat v oblasti sociálních sítí.

2 Analytické metody využité při analýze WoM

Analýza byla provedena pomocí technologie Autonomy IDOL¹ poskytující infrastrukturu pro řízení a procesování větších objemů nestrukturovaných dat. Data nejsou uložena relačně, ale v proprietární dokumentové struktuře a optimalizována pro rychlé zpracování. IDOL dokáže kontextuálně rozumět obsahu dat nezávisle na formátu. Pomocí statistických metod, dominujících nezávisle nad metodami přirozeného zpracovávání jazyka, provádí s daty analytické operace. Jádrem je identifikace vzorů, které se přirozeně vyskytují v textu na základě jejich užití a výskytu, které za určitých podmínek korespondují s daným kontextem. Technologie si buduje svou vlastní pravděpodobnostní mapu různých termínů, dle níž poté vrací relevantní výsledky na uživatelsky spuštěné operace, funkce a dotazy.

Datový vzorek byl získán ze sociální sítě Facebook přes Autonomy konektory pro sociální média. Obsahuje veškeré veřejně dostupné příspěvky z veřejných stránek českých bankovních institucí za dobu jednoho roku od dubna 2013 do března 2014. Celkově 28 289 dokumentů od průměrně 200 uživatelů měsíčně se skládá v názvosloví Facebooku z 22 629 komentářů, 4 542 feedů, 1 061 fotografií a 746 postů.

Najít užitečné informace v takových datech není triviální problém, musí se postupovat kaskádově po kouscích informací vydolovaných z textu. V takových situacích vyvstávají otázky, jako zda bude informace užitečná, či jak by měla vypadat, aby užitečná byla. Cílem bylo vyhnout se vymyšlení případů, které by mohly, ale také nemusely nastat. Je třeba také zmínit, že analýza nebere v potaz cenzury textů (mazání nevhodných příspěvků) ze strany administrátorů dané stránky, což je z hlediska vyhodnocování třetí stranou nerealné. Pokud by tímto vyhodnocovatelem byla sama banka, je možné zohlednit její vlastní cenzuru ovšem nikoliv konkurence.

Analýza vycházela z výzkumu provedeným Harvard Business Review v roce 2010 (H.b., 2010) zkoumající různé aspekty a problematiku vztahů firmy a sociálních sítí. Část výzkumu je zaměřena na tematiku primárních benefitů společností využívajících sociální sítě a jejich

¹ Technologie zakoupená společností HP. IDOL = Intelligent Data Operating Layer.

práce s nimi. Ukazuje, které primární důvody pro využívání sítí převládají v efektivních společnostech:

- Zvyšování povědomí o organizaci, značce, jejích produktech a službách mezi cílovými zákazníky.
- Zvyšování pozitivnějšího vnímání organizace a jejích produktů.
- Schopnost monitorovat, co bylo o organizaci řečeno. (H.b., 2010)

Pokud budeme uvažovat bankovní instituce jako efektivní byznys uživatele sociálních sítí, závěry analýzy musí zodpovědět přinejmenším otázku, jak si organizace vede ve třech oblastech zmíněných výše. Na základě analytických výsledků by potom měly být spuštěny procesy vedoucí ke spokojenějším a loajálním zákazníkům.

Povědomí bylo analyzováno srovnáním vývoje počtu příspěvků na sociálních stránkách jednotlivých institucí a identifikací počtu unikátních přispěvatelů v jednotlivých měsících.

Jelikož datový vzorek obsahoval velké množství příspěvků, jejichž obsah nebyl znám, bylo využito shlukování jako nesupervizované techniky Machine Learningu. Ta se používá za podmínek, kdy není určena třída, se kterou by byl výstup asociován, ale vstupní instance mohou být rozděleny do přirozených skupin – shluků, které jsou navzájem homogenní ale mezi sebou heterogenní. Literatura popisuje mnoho algoritmů shlukování, např. (Feldman a Sanger, 2007; Berka, 2003; Witten at al., 2011; Tsitsis a Chorianopoulos, 2009; Kruengkrai a Jaruskulchai, 2002). Nestrukturovaná data vstupují do shlukování jako celé dokumenty, kde jsou data pro lepší přístup uložena v předdefinované struktuře, v tomto případě v XML a IDX² formátu. Dokumenty jsou rozděleny do individuálních shluků reprezentujících konceptuální prostor se znalostní bází obsahující sadu položek se společnými charakteristikami. Základním předpokladem je hypotéza, že relevantní dokumenty mají tendenci si být navzájem více podobné než ty irelevantní, a tak pomáhají zlepšovat efektivitu vyhledávání, pokud je omezeno pouze na relevantní dokumenty. Postup analýzy byl proveden shora-dolů, z velkých jednotek shluků jsme tvořili pod-shluky, které zúžily data na jednotlivá témata. Nad těmito shluky jsme pak tvořili statistiky, dle kterých byl dodatečně dohledáván kontext, který objasňuje, *co bylo o dané bance řečeno*.

K objasnění významu jednotlivých shluků byly příspěvky rozděleny do tří skupin dle sentimentu, jenž byl předem napočítán:

- Příspěvky spjaté s negativním sentimentem,
- příspěvky spjaté s pozitivním sentimentem,
- všechny příspěvky nehledě na sentiment.

Analýza sentimentu pak dává také odpověď na otázku, *jak je banka vnímána* jejími zákazníky či prospekty. Kromě evaluace sentimentu je možné v datech číst také počty „líbí se mi“ pod každým příspěvkem a jejich součty za dané banky porovnávat v čase.

Sentiment analýza využívá množství algoritmů, některé popisuje (Liu, 2008). V tomto výzkumu bylo využito klasifikační metody, tzv. opinion classification, jež využívá polarity sentimentu klasifikující, zda daný dokument vyjadřuje pozitivní nebo negativní o určitém objektu. (Kaur et al., 2013) V této analýze byl započítáván také neutrální sentiment, jehož váha je rovna nule a bipolární sentiment nesoucí negativní a pozitivní sentiment zároveň. Bipolární klasifikace se ale těžko určuje na jakémkoliv měřítku. V našem případě má daný termín skóre v negativním (záporná váha) i pozitivním případě (kladná váha), ale výsledek a přesnost musí být odvozena z datového vzorku. Váhy sentimentu jednotlivých slov jsou již přednastaveny v knihovně IDOL ve slovnících pro jednotlivé jazyky, sentiment chybějících

² XML = Extensible Markup Language a IDX (proprietární Autonomy formát) jsou typy značkovacího jazyka, které strukturují daný soubor.

slov lze ručně nadefinovat. Váha sentimentu je počítána jako souhrnné (absolutní) skóre sentimentu vyděleného součtem (rozdílem pro absolutní skóre) počtu negativního a pozitivního výskytu sentimentu. Výsledkem algoritmu je pak příznak s typem sentimentu a jeho hodnotou. Poté, co je spočítán sentiment nad příspěvkem, slovník vypíše polaritu a váhu sentimentu každého ze slov v příspěvku, vyhodnotí sentiment celého řetězce jako celku a určí téma (předmět), které tuto polaritu determinuje.

```
function handler(document)

    positive = 0
    negative = 0
    score = 0
    type = 0
    for ii,field in ipairs { document:findField("*") } do
        local fieldname = document:fieldName(field)
        --print(fieldname)
        if fieldname == "SENTIMENT_POSITIVE" then
            positive = positive + 1
        elseif fieldname == "SENTIMENT_NEGATIVE" then
            negative = negative + 1
        elseif fieldname == "SCORE" then
            if type == 1 then
                score = score + document:fieldGetValue(field)
            elseif type == -1 then
                score = score - document:fieldGetValue(field)
            end
        end
    end

    document:addField("SENTIMENT_POSITIVE_COUNT", positive)
    document:addField("SENTIMENT_NEGATIVE_COUNT", negative)
    document:addField("SENTIMENT_COUNT", positive - negative)
    document:addField("SENTIMENT_SCORE", score)
    document:addField("SENTIMENT_WEIGHT", score / (positive + negative))
    document:addField("SENTIMENT_WEIGHT2", math.abs(score) / (positive - negative))

end
```

Tabulka 1: Kód edukčního modulu pro počítání sentimentu. Zdroj (Autonomy, 2013).

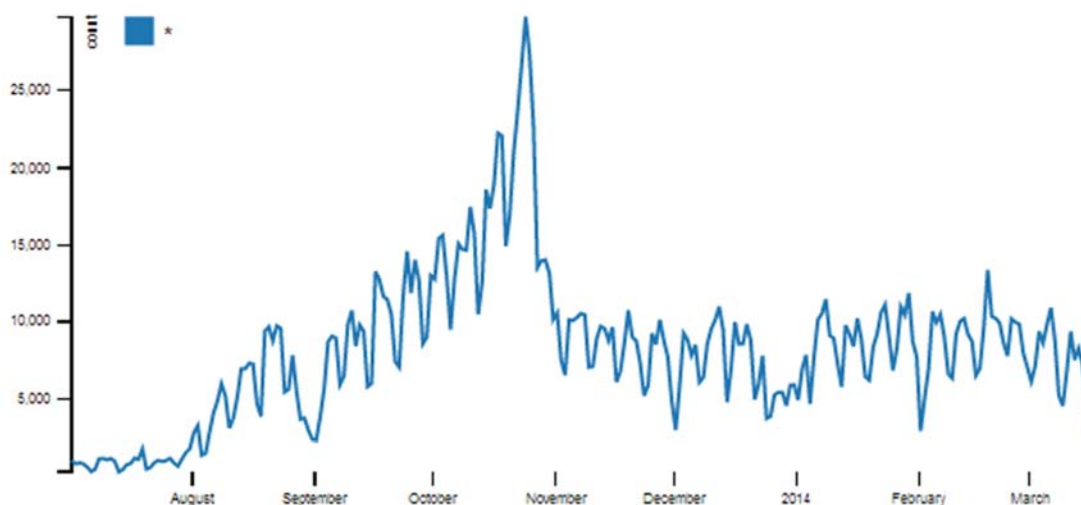
3 Průběh a výsledky analýzy dat z Facebooku

Z identifikace společných vrcholů počtu příspěvků jednotlivých bank (Graf 2) lze pozorovat, kdy všechny banky vložily více příspěvků než předtím a kdy lze pozorovat významný pokles, který v tomto případě následuje vždy po vysokém růstu. Interpretovat to lze skutečností, že v určitý čas začnou banky promovat určitý produkt, nějakou událost nebo řešit nějaký problém aktivně s uživateli. Pokud je problém vyřešen, událost se již konala nebo nemá smysl dále promovat, počet příspěvků začne silně klesat. Z celkového počtu příspěvků měsíčně (Graf 1) byl vypořizován největší nárůst v říjnu 2013 (následovaný opětovným poklesem).

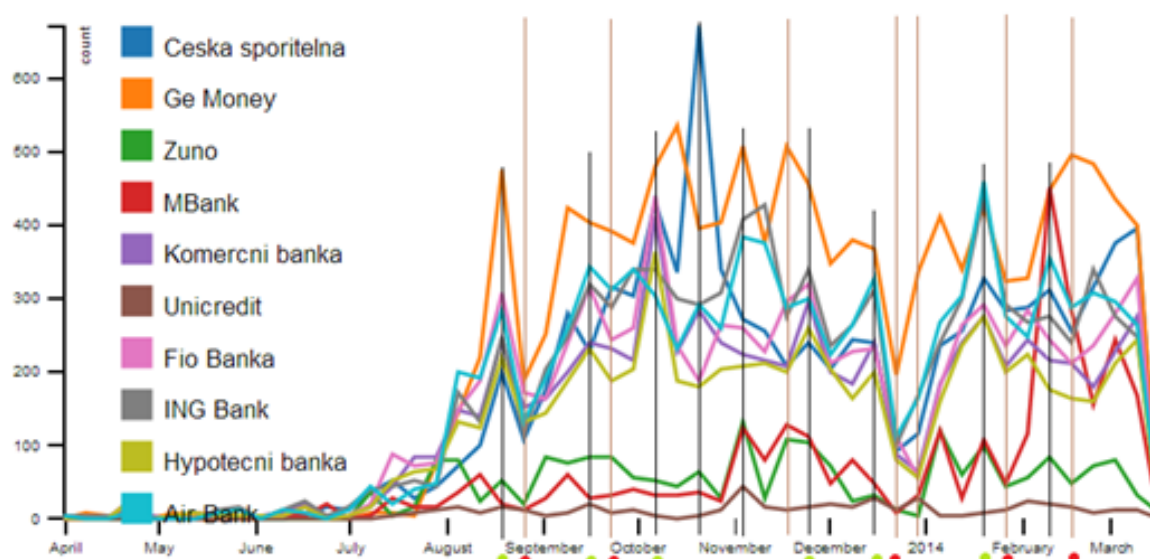
Z analýzy kontextu a sentimentu (Graf 4, 5 a Tabulka 2) bylo vyvozeno, že tento nárůst byl způsoben vyřešením některých problémů spojených s internetovým bankovníctvím u Komerční banky. Na začátku října Komerční banka upravila svou aplikaci pro internetové

bankovníctví MojeBanka bez nutnosti běhu uživateli neoblíbené Javy. Pozornosti se také dočkalo otevření nové pobočky na Smíchově. Na konci října se také výrazně zvýšil počet příspěvků České spořitelny. Kontextová analýza ukázala, že nárůst způsobilo uvedení kampaně, kdy Česká spořitelna ke každému novému účtu dávala kávovar jako dárek, což způsobilo mnoho jak pozitivních, tak negativních uživatelských ohlasů. Česká spořitelna byla aktivní také v uvedení dalších programů jako zvyšování finanční gramotnosti mezi dětmi nebo navrácení uměleckých exhibic do galerie České spořitelny.

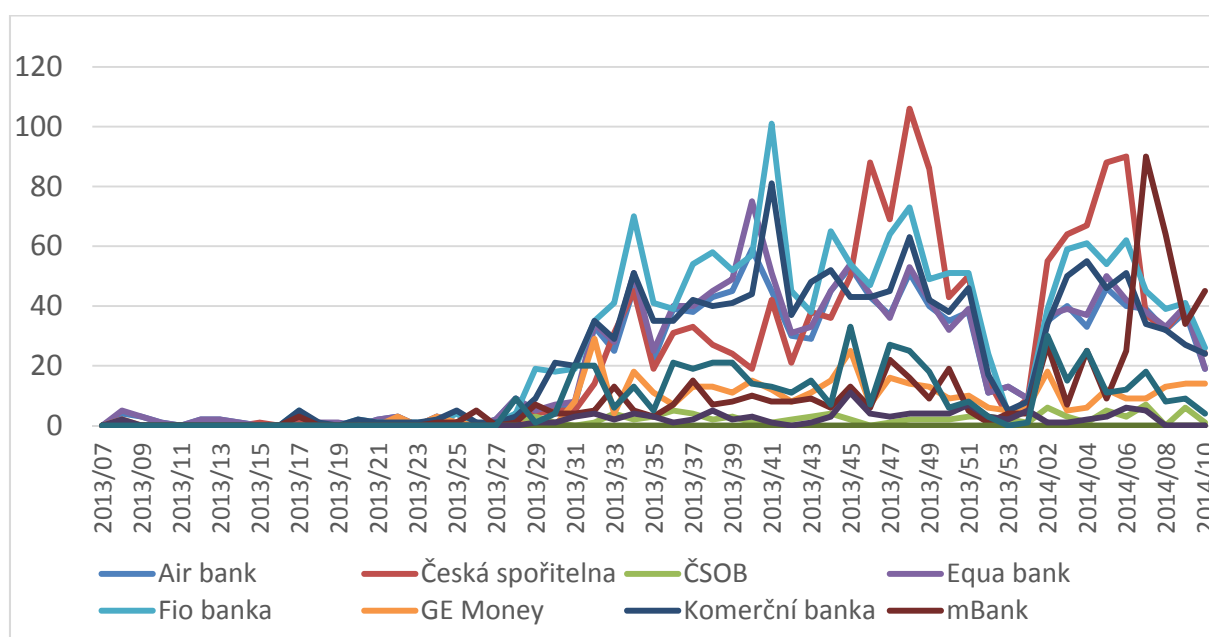
Graf 3 srovnává počty příspěvků obsahujících název banky v součtech po týdnech. Zajímavé je odmlčení na konci roku, kdy si pravděpodobně uživatelé nechťeli kazit vánoční svátky psaním příspěvků o bankách. Oproti tomu lze sledovat vysoké nárůsty v příspěvcích zmiňujících Fio banku na konci září. Na základě shluku provedeným nad tímto obdobím bylo zjištěno, že banka spustila akci „karta Fio banky s Jawou kolem světa“. Ve 40. týdnů lze z množství příspěvků České spořitelny vyčíst uvedení a silné propagování loajálního programu IBod, odstartovaného 5. října. V prvních šesti týdnech se objevil IBod v příspěvcích průměrně osmkrát. Program byl vnímán velmi pozitivně. Nárůst příspěvků mBank na začátku roku 2014 způsobil vznik a prezentace nového internetového bankovníctví. Zajímavé jsou nízké hodnoty ČSOB. Tato banka není mezi autory vlastních příspěvků, proto také není mezi monitorovanými bankami. Naměřené hodnoty byly získány skrz zmínky na stránkách jiných bank. Naopak Unicredit banka sama vytvořila 587 postů jako sedmá nejvíce aktivní banka, ale její křivka je pouze o něco výše než ČSOB, protože není často zmiňována v obsahu příspěvků a lidé o ní tolik nemluví.



Graf 1: Celkový počet příspěvků měsíčně. Zdroj: Autorka.

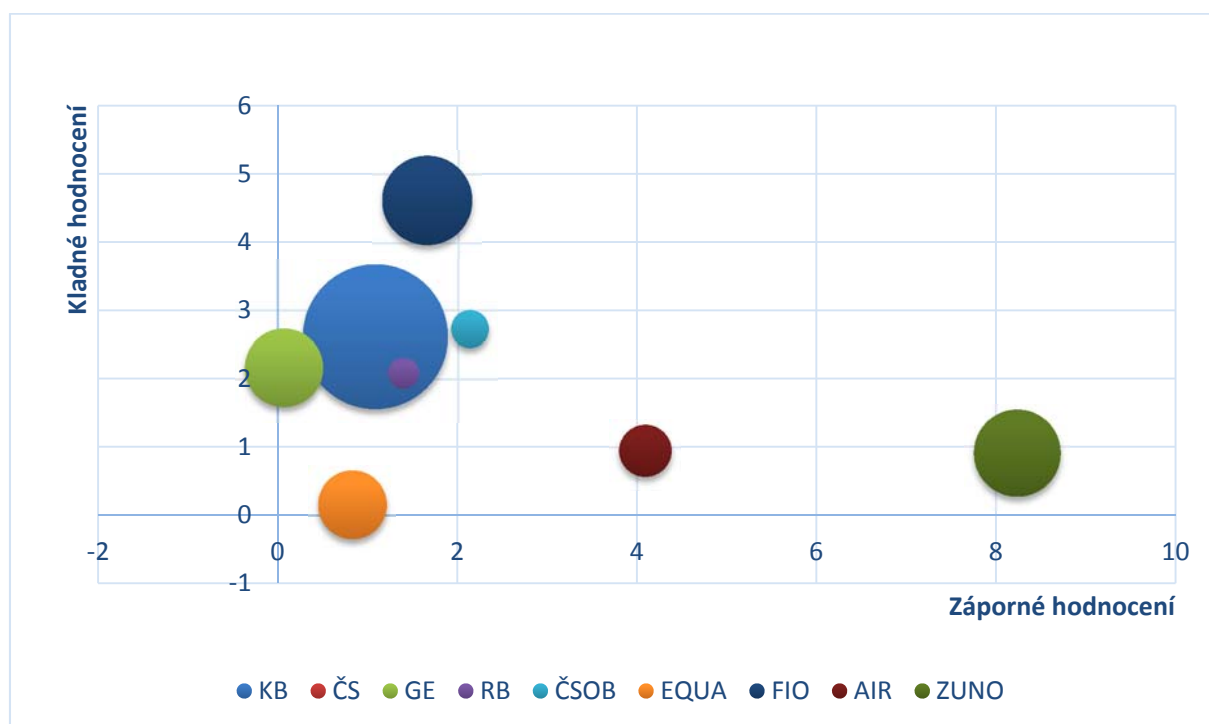


Graf 2: Počet příspěvků bank měsíčně. Zdroj: (Sperkova, 2014).



Graf 3: Počet uživatelských příspěvků nesoucí název banky týdně. Zdroj: (Sperkova, 2014).

Dle výpočtu sentimentu popisovaným Tabulkou 1 vznikl Graf 4. Pokud je vyznění příspěvků spíše pozitivní, jsou bubliny více vlevo nahoře, naopak negativní jsou vpravo dole. Velikost bubliny ukazuje, kolik příspěvků daný název banky obsahoval. Sentiment analýza ukázala, že nejpozitivněji uživatelé mluví o Fio Bance, kdežto nejzápornější hodnocení má Zuno. Obě banky jsou zmiňovány ve stejném množství příspěvků. Nejméně příspěvků obsahuje název Raiffeisenbank, důvodem může být její malá aktivita na síti, je autorem pouze patnácti postů. Jiný důvod může být, že v analýze jsou brány v potaz pouze oficiální názvy bank, nikoliv mutace vytvořené uživateli, jako např. „Raiffka“. Nejvíce záznamů obsahuje Komerční banka s relativně pozitivním výsledkem.



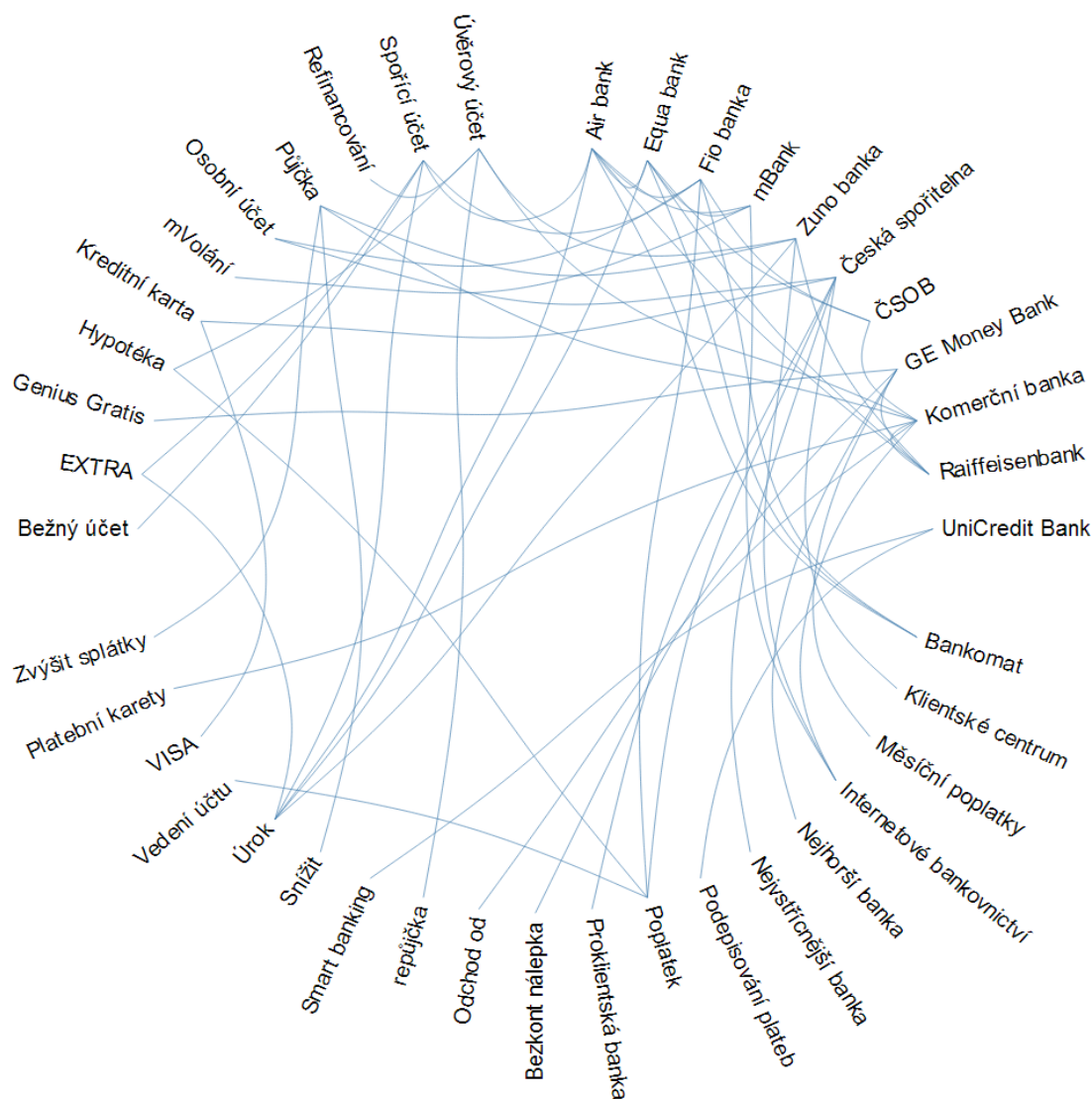
Graf 3: Sentiment analýza příspěvků obsahujících název banky. Zdroj: (Šperková, 2014)

Přes celou finanční oblast vzniklo 6 shluků, z nichž těmi největšími byly: Online bankovníctví/internet banking, Platební karty a Genius gratis. Přes klíčová slova v těchto shlucích byly vytvořeny pod-shluky (Tabulka 2), jejichž počet je udán v závorce za výrazem a ty největší jsou uvedeny pod výrazy.

• „úrok“ (12)
• spořicí účet
• „kreditní“ (1)
• Kreditní karta / debetní karta
• „poplatky“ (1)
• Studentský účet
• „úvěr“ (1)
• Anuitní, hypoteční, splacené
• „Air“ (9)
• srovnání s dalšími bankami
• úrokové sazby na vkladových účtech
• dostupnost online služeb
• „Zuno“ (1)
• srovnání s dalšími „novými“ bankami
• „Equa“ (2)
• srovnání s dalšími „novými“ bankami
• spořicí účet
• „GE Money“ (2)
• Genius Gratis
• před-nabité bezkontaktní nálepky
• „Komerční banka“ (3)
• internetové bankovníctví
• mobilní banka
• refinancování
• „ČSOB“ (4)
• srovnání s dalšími bankami
• bankomat ČSOB
• „Česká spořitelna“ (1)
• klientské centrum, péče o klienty

Tabulka 2: Seznam pod-shluků vytvořených nad klíčovými slovy. Zdroj: Autorka

S využitím JavaScript knihovny, vyhledávacích slov a výsledků ze shlukování byla vytvořena bundling mapa (Graf 5) obsahující vztahy mezi tématy, které se často spolu vyskytují. Nutné bylo data pročistit, aby v mapě zůstaly jen silné pojmy, tedy ty, které se vyskytovaly nejčastěji a zároveň měly silné vazby na pojmy další. Například pojem „re-půjčka“ se sice vyskytuje ve 104 příspěvcích, ale jeho další vazby nejsou nejspíše natolik silné, aby dávalo smysl je zobrazovat, proto mu zůstala vazba pouze na pojem „úvěrový účet“. Lze vidět, že názvy bank se často vyskytují spolu, jak lidé mají tendence služby bank porovnávat. Která banka si v povědomí lidí vede lépe je vidět v bublinovém Grafu 4.



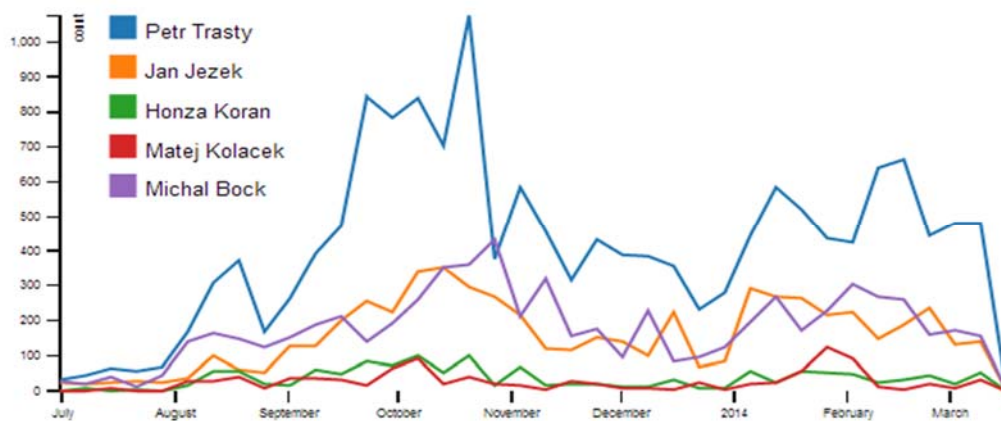
Graf 4: Mapa vazeb mezi tématy. Zdroj: (Šperková, 2014)

Jelikož na Facebooku nelze přispívat anonymně, jsou tak známy autoři všech příspěvků. Graf 6 ukazuje nejoblíbenější stránky (líbí se mi) uživatelů, kteří na stránkách bank nechali 50 a více příspěvků a zároveň přispívali v říjnu 2013, který byl vyhodnocen jako nejaktivnější měsíc. Zajímavostí je, že z těchto asi 300 uživatelů byla jen zhruba třetina žen. Z těchto výsledků lze odvodit, že se nejvíce lidí zajímá kromě bank také o telekomunikace, seriály či řetězec Lidl. Velikost obdélníku znamená vyšší počet lidí. Z výsledku této analýzy vyplývá doporučení spolupráce s jinými firmami, o které se zákazníci zajímají nejvíce, např. vytvořením společného produktu, darování poukázek na produkty druhé firmy, vytvořením společných akcí apod. Příkladem je akce České spořitelny, která ke každému novému účtu darovala kávovar firmy, se kterou kooperuje.

Z analýzy bylo dále identifikováno pět nejaktivnějších autorů a jejich aktivita v čase (Graf 7). Analýzou uživatelských „líbí se mi“ jsme zjistili, že tito uživatelé jsou aktivní také v jiných oblastech než finance, jako telekomunikace, politika nebo IT. Vzhledem k jejich aktivitě na síti tak mohou mít vysokou autoritu mezi dalšími uživateli jako tzv. opinion makers. Druhou možností je, že jsou to bankami najatí uživatelé, jejichž prací je právě přispívat na jejich stránkách a ovlivňovat tak zákazníky.



Graf 5: Počty "líbí se mi" neaktivnějších uživatelů za říjen 2013. Zdroj: Autorka



Graf 6: Identifikace neaktivnějších uživatelů dle součtu jejich příspěvků. Zdroj: (Sperkova, 2014)

4 Diskuze a závěr

Ačkoliv výzkum provedený Harvard Business Review (2010) dospěl k názorům, že společnosti pracující efektivně se sociálními sítěmi jsou mimo jiné také více zaměřeny na zlepšování vnímání firmy zákazníkem, musí být poznamenáno, že jakékoliv úsilí v tomto

smyslu má svá omezení, která nemohou být ignorována. Je třeba si uvědomit, že v bankovníctví se jedná pouze o jeden z kanálů, se kterými zákazník přichází do styku s bankou. Jenom existence samotných sociálních sítí neznamená konec tradičních kanálů jako bankovních přepážek či call center nebo omezení jejich významu a vhodnosti. Lze ale tvrdit, že na rozdíl od těchto kanálů je Facebook veřejný, jedinečný obousměrný komunikační kanál, revoluční ve své bezprostřednosti a blízkosti bank svým zákazníkům. Obsah vidí daleko více zákazníků či prospektů, tedy tak může regulovat jejich názor na danou banku více než jiné kanály. Facebook tak má významný vliv na vytváření všeobecného vnímání zákazníky. Banka se může buď veřejně ztrapňovat, nebo ukázat, že je svým zákazníkům nápomocná. Přesto je těžké určit, do jaké míry je reputace banky je měněna skrz její aktivity na síti a jak vysoký podíl na tom má reálná hodnota jejich produktů a služeb. Nicméně data na Facebooku mohou reflektovat vliv banky na své zákazníky jak skrz síť, tak skrz jiné kanály. Tento vztah mezi kanály a reputací banky je M:1.

Bankovní sektor je na rozdíl od retailového sektoru nabízejícího spotřební zboží specifický i lehce rozdílným chováním svých uživatelů na sociálních sítích. Lze vyvodit, že lidé využívají síť bankovních institucí relativně aktivně, pokud jde o nějaký problém, např. s internetovým bankovníctvím, výši poplatků za vedení účtů, dále diskutují názory na kampaně vedené bankou a srovnávají banky mezi sebou. V datech také bylo často zaznamenáno, že dle názorů uživatelů, banky nejsou schopny řešit problémy svých zákazníků skrz návštěvu na pobočce nebo volání na call-centrum. Využívají tak Facebook jako technickou podporu. Ostatní lidé reagují na problémy druhých celkem aktivně a bezprostředně, proto je nutné soustředit pozornost na vhodné metody, jak reagovat na tyto stížnosti. To je nutné uvážit i při definici profilu zaměstnanců, kteří řídí bankovní sociální síť, jejich přípravě a školení.

Doporučení pro banky je přispívat, chovat se aktivně a proaktivně tak, aby měl koncový zákazník dojem, že i přestože má banka problém, je schopna ho řešit, což se nyní v mnoha případech neděje. Produkty by měly být promovány kontinuálně, nikoliv nárazově, a zároveň interaktivní cestou. I když na druhou stranu je nutné říct, že kontinuita je v současné době v praxi nereálná zejména z finančních a procesních důvodů. Musel by se tedy klást také důraz na změnu v těchto procesech. V dnešní spotřebitelské době nemají lidé problém se nahlas ozvat, když se jim něco nelíbí a je nutné, aby na tato volání byli poskytovatelé schopni flexibilně reagovat.

Poděkování:

Tento příspěvek byl vytvořen díky podpoře z grantu IGA F4/18/2014 řešeném na Fakultě informatiky a statistiky, VŠE v Praze.

Seznam použitých zdrojů

- Barbosa, L., & Feng, J. (2010, August). Robust sentiment detection on twitter from biased and noisy data. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters* (pp. 36-44). Association for Computational Linguistics.
- Berka, P. (2003). *Dobývání znalostí z databází*. 1. vydání. Praha: Academia, 366 s. ISBN 80-200-1062-9.
- Bermingham, A., Conway, M., McInerney, L., O'Hare, N., & Smeaton, A. F. (2009, July). Combining social network analysis and sentiment analysis to explore the potential for online radicalisation. In *Social Network Analysis and Mining*, 2009. ASONAM'09. International Conference on Advances in (pp. 231-236). IEEE.
- Bughin, J., Doogan, J., & Vetvik, O. J. (2010). A new way to measure word-of-mouth marketing. *McKinsey Quarterly*, 2, 113-116.

- ČNB. (online 11. 6. 2014) Seznam měnových finančních institucí v České republice. [cit: 2014-11-06]
Retrieved from ČNB:
http://www.cnb.cz/miranda2/export/sites/www.cnb.cz/cs/statistika/menova_bankovni_stat/seznamy_mbs/s_eznam_mfi/CZ_MFI_CZ_verze.xls
- Feldman, R., & Sanger, J. (Eds.). (2007). *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge University Press.
- Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*, 1-12.
- Griffin, A., & Hauser, J. R. (1993). The voice of the customer. *Marketing science*, 12(1), 1-27.
- H.b. (01. 08. 2010). The New Conversation: Taking Social Media from Talk to Action. [online] 24 pages. *Harvard Business Review*, 10815-PDF-ENG. Retrieved from http://www.sas.com/resources/whitepaper/wp_23348.pdf
- Han, X., & Niu, L. (2012). Word of mouth propagation in online social networks. *Journal of Networks*, 7(10), 1670-1676.
- Helm, S., & Schlei, J. (1998). Referral potential–potential referrals. An investigation into customers' communication in service markets. In *Track 1 Market Relationships, Proceedings 27th EMAC Conference, Marketing Research and Practice* (pp. 41-56). ISQA.
- Hu, N., Pavlou, P. A., & Zhang, J. (2006, April). Can online word-of-mouth communication reveal true product quality? Experimental insights, econometric results, and analytical modeling. In *Proceedings of the 7th ACM Conference on Electronic Commerce*, Ann Arbor, Michigan, USA, June.
- Choi, J. H., & Scott, J. E. (2013). Electronic word of mouth and knowledge sharing on social network sites: a social capital perspective. *Journal of theoretical and applied electronic commerce research* 8(1), 69-82.
- Kaur, H. (2013). Opinion Mining Task and Techniques: A Survey. *International Journal of Advanced Research in Computer Science*, 4(3).
- Kruengkrai, C., & Jaruskulchai, C. (2002, July). A parallel learning algorithm for text classification. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 201-206).
- Liu, Bing. (2008) Web data mining. Exploring hyperlinks, contents, and usage data, Corr. 2. print. *Data-centric systems and applications*. Springer, Berlin.
- Morrissey, S. (2012). Digitální svět nové generace mění bankovníctví. In *Bankovníctví v proměnách času: Sborník textů k 20 letům České bankovní asociace*. Praha: ČBA, s. 72-75. ISBN 978-80-260-3016-4.
- Nacher, P. (online 30.9.2014) Druhá fáze přechodu bankovních klientů se blíží. *Bankovní poplatky*. Retrieved from <http://www.bankovnipoplatky.com/druha-faze-prechodu-bankovnich-klientu-se-blizi-25121>
- Pak, A., & Paroubek, P. (2010, May). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In *7th International Conference on Language Resources and Evaluation*, (pp. 1320-1326).
- Půlpánová S. (2007). *Komerční bankovníctví v České republice*. 1. vyd. Praha: Oeconomica, 338 s. ISBN 978-80-245-1180-1.
- Sperkova, L. (2014) Word of Mouth Analysis on Facebook in Banking. In *Marketing Identity: Explózia inovácií*. Trnava.
- Tsiptsis, K., & Chorianopoulos, A. (2011). *Data mining techniques in CRM: inside customer segmentation*. John Wiley & Sons.
- Veeck, A., & Hoger, B. (2014). Tools for Monitoring Social Media: A Marketing Research Project. *Marketing Education Review*, 24(1), 37-72.
- Witten, I. H., Frank, E., & Mark, A. Hall (2011). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.