

Univerzita Pardubice

Fakulta ekonomicko-správní

Vliv emocí v novinových článcích na abnormální akciové výnosy

Věra Zdražilová

**Diplomová práce
2019**

ZADÁNÍ DIPLOMOVÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Bc. Věra Zdražilová**
Osobní číslo: **E16629**
Studijní program: **N6208 Ekonomika a management**
Studijní obor: **Ekonomika a management podniku**
Název tématu: **Vliv emocí v novinových článcích na abnormální akciové výnosy**
Zadávající katedra: **Ústav podnikové ekonomiky a managementu**

Z á s a d y p r o v y p r a c o v á n í :

Záměrem práce je shrnout současné přístupy k predikci abnormálních akciových výnosů, charakterizovat přístupy k analýze emocí v textových dokumentech, provést sběr dat z novinových serverů, provést analýzu emocí v těchto datech a použít tato data k predikci abnormálních akciových výnosů.

Osnova:

- Současné přístupy k predikci abnormálních akciových výnosů
- Současné přístupy k analýze emocí v textových dokumentech
- Sběr a zpracování dat
- Aplikace vhodného postupu pro analýzu emocí v těchto datech a jejich použití k predikci abnormálních akciových výnosů
- Interpretace získaných výsledků

Rozsah grafických prací:

Rozsah pracovní zprávy: cca 50 stran

Forma zpracování diplomové práce: tištěná/elektronická

Seznam odborné literatury:

HENDL, Jan. Kvalitativní výzkum: základní teorie, metody a aplikace. 2., aktualiz. vyd. Praha: Portál, 2008. ISBN 978-80-7367-485-4.

LIU, Bing. Sentiment analysis and opinion mining. San Rafael: Morgan & Claypool Publishers, 2012. ISBN 978-1608458844.

PETERSON, Richard L. Trading on sentiment: the power of minds over markets. Hoboken, New Jersey: Wiley, 2016. ISBN 978-1-119-12276-0.


VALACH, Josef. Investiční rozhodování a dlouhodobé financování. 3., přeprac. a rozš. vyd. Praha: Ekopress, 2010. ISBN 978-80-86929-71-2.

Vedoucí diplomové práce: doc. Ing. Petr Hájek, Ph.D.


Ústav systémového inženýrství a informatiky

Datum zadání diplomové práce: 3. září 2018

Termín odevzdání diplomové práce: 30. dubna 2019


doc. Ing. Romana Provozničková, Ph.D.
děkanka

L.S.


doc. Ing. Marcela Kožená, Ph.D.
vedoucí ústavu

V Pardubicích dne 3. září 2018

PROHLÁŠENÍ

Prohlašuji, že jsem tuto práci vypracovala samostatně. Veškeré literární prameny a informace, které jsem v práci využila, jsou uvedeny v seznamu použité literatury.

Byla jsem seznámena s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorský zákon, zejména se skutečností, že Univerzita Pardubice má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona, a s tím, že pokud dojde k užití této práce mnou nebo bude poskytnuta licence o užití jinému subjektu, je Univerzita Pardubice oprávněna ode mne požadovat přiměřený příspěvek na úhradu nákladů, které na vytvoření díla vynaložila, a to podle okolností až do jejich skutečné výše.

Beru na vědomí, že v souladu s § 47b zákona č. 111/1998 Sb., o vysokých školách a o změně a doplnění dalších zákonů (zákon o vysokých školách), ve znění pozdějších předpisů, a směrnicí Univerzity Pardubice č. 9/2012, bude práce zveřejněna v Univerzitní knihovně a prostřednictvím Digitální knihovny Univerzity Pardubice.

V Pardubicích dne 30.6.2019

Věra Zdražilová

PODĚKOVÁNÍ:

Tímto bych ráda poděkovala především svému vedoucímu práce doc. Ing. Petru Hájkovi, Ph.D. za jeho odbornou pomoc, cenné rady a poskytnuté materiály, které mi pomohly při zpracování diplomové práce.

Děkuji také své rodině a blízkým za veškerou podporu a pomoc, kterou mi poskytovali nejen při psaní mé diplomové práce, ale i během celého mého studia.

ANOTACE

Tato diplomová práce je zaměřena na predikci abnormálních akciových výnosů za pomoci analýzy sentimentu. Po teoretickém shrnutí současných přístupů k predikci akciových výnosů je vysvětlena samotná analýza sentimentu. Následně je proveden sběr a zpracování dat k analýze sentimentu z novinových článků. Pomocí regresní analýzy je posouzena závislost mezi sentimentem vyjádřeným v novinových článcích a vývojem cen akcií na kapitálových trzích.

KLÍČOVÁ SLOVA

Analýza sentimentu, emoce, slovník sentimentu, akcie, abnormální akciové výnosy, regresní analýza

TITLE

Impact of Emotions in Newspaper Articles on Abnormal Stock Returns

ANNOTATION

This thesis is focused on prediction of abnormal stock returns using sentiment analysis. After the theoretical summary of the current approaches to the prediction of stock returns, the sentiment analysis itself is explained. Subsequently, data collection and processing is performed to sentiment analysis from newspaper articles. The regression analysis assesses the dependence between sentiment expressed in newspaper articles and the evolution of stock prices on the capital markets.

KEYWORDS

Sentiment Analysis, Emotions, Sentiment Lexicon, Stocks, Abnormal Stock Returns, Regression Analysis

OBSAH

ÚVOD	10
1 AKCIOVÉ TRHY	12
1.1 AKCIE	12
1.1.1 <i>Druhy akcií</i>	13
1.1.2 <i>Burza cenných papírů</i>	15
1.1.3 <i>Burzovní index</i>	16
1.2 ABNORMÁLNÍ AKCIOVÝ VÝNOS	16
1.3 SOUČASNÉ PŘÍSTUPY K PREDIKCI ABNORMÁLNÍCH AKCIOVÝCH VÝNOSŮ	17
1.3.1 <i>Fundamentální analýza</i>	17
1.3.2 <i>Technická analýza</i>	21
1.3.3 <i>Psychologická analýza</i>	24
2 ANALÝZA SENTIMENTU	27
2.1 DEFINICE	27
2.2 NÁZOR	28
2.2.1 <i>Druhy názorů</i>	28
2.2.2 <i>Prvky názoru</i>	29
2.3 ÚROVNĚ ANALÝZY SENTIMENTU.....	31
2.4 METODY ANALÝZY SENTIMENTU	32
2.4.1 <i>Strojové učení</i>	32
2.4.2 <i>Slovníkové metody</i>	32
2.5 METODOLOGIE.....	33
2.5.1 <i>Sběr dat</i>	33
2.5.2 <i>Příprava textu</i>	34
2.5.3 <i>Detekce sentimentu</i>	34
2.5.4 <i>Klasifikace sentimentu</i>	35
2.5.5 <i>Prezentace výstupů</i>	35
2.6 PROBLÉMY ANALÝZY SENTIMENTU.....	35
2.6.1 <i>Subjektivita názoru</i>	36
2.6.2 <i>Sarkasmus a ironie</i>	36
2.6.3 <i>Falešné recenze</i>	36
2.6.4 <i>Problémy slovníků</i>	37
3 ANALÝZA SENTIMENTU V NOVINOVÝCH ČLÁNCÍCH	38
3.1 SBĚR A PŘÍPRAVA DAT.....	38
3.2 MĚŘENÍ SENTIMENTU	41
3.3 VÝSLEDKY MĚŘENÍ SENTIMENTU	47
4 VLIV EMOCÍ NA ABNORMÁLNÍ VÝNOS AKCIE	56
4.1 VÝPOČET ABNORMÁLNÍHO VÝNOSU AKCIE.....	56
4.2 REGRESNÍ ANALÝZA.....	57
ZÁVĚR	60
POUŽITÁ LITERATURA	62
SEZNAM PŘÍLOH	65

SEZNAM TABULEK

Tabulka 1 - Seznam vybraných firem	39
Tabulka 2 - Výsledný počet dní	40
Tabulka 3 - Průměr naměřeného sentimentu za sledované období	48

SEZNAM ILUSTRACÍ

Obrázek 1 - Vnitřní hodnota akcie	15
Obrázek 2 - Primární, sekundární a terciální trend - teorie Charles Dow	22
Obrázek 3 - Základní rámec Elliottových vln	24
Obrázek 4 - Ukázka textového souboru	40
Obrázek 5 - Nejčastější pozitivní slova dle Liu	42
Obrázek 6 - Nejčastější pozitivní slova dle Loughrana a McDonalda	42
Obrázek 7 - Nejčastější pozitivní slova dle Henry	43
Obrázek 8 - Nejčastější negativní slova dle Liu	44
Obrázek 9 - Nejčastější negativní slova dle LM	44
Obrázek 10 - Nejčastější negativní slova dle Henry	45
Obrázek 11 - Příklad problému slovníku sentimentu	46
Obrázek 12 - Počty pozitivních a negativních slov dle jednotlivých slovníků	46
Obrázek 13 - Poměry naměřeného sentimentu	47
Obrázek 14 - Průměrné měsíční hodnoty sentimentu firmy AAPL	49
Obrázek 15 - Průměrné měsíční hodnoty sentimentu firmy AMZN	49
Obrázek 16 - Průměrné měsíční hodnoty sentimentu firmy JPM	50
Obrázek 17 - Průměrné měsíční hodnoty sentimentu firmy XOM	51
Obrázek 18 - Průměrné měsíční hodnoty sentimentu firmy WFC	52
Obrázek 19 - Průměrné měsíční hodnoty sentimentu firmy GE	52
Obrázek 20 - Průměrné měsíční hodnoty sentimentu firmy WMT	53
Obrázek 21 - Průměrné měsíční hodnoty sentimentu firmy T	54
Obrázek 22 - Průměrné měsíční hodnoty sentimentu firmy PFE	54
Obrázek 23 - Průměrné měsíční hodnoty sentimentu firmy MSFT	55
Obrázek 24 - Ukázka modelu logistické regrese pro firmu AAPL	59

SEZNAM ZKRATEK A ZNAČEK

a.s.	akciová společnost
CP	cenné papíry
L&M	Loughran a McDonald
OTC	over-the-counter
USD	americký dolar

ÚVOD

Každý den vychází tisíce novinových článků, které přináší nejnovější události ze všech možných oblastí, a co možná nejrychleji o nich informují širokou veřejnost po celém světě. Díky prudkému rozvoji internetu klesá zájem o klasické tištěné noviny, jež byly doposud zdrojem novinových článků a dochází tak k přesunu čtenářů na internetová média. Lidé tak mají nejrychlejší a nejsnadnější přístup k informacím v celé historii. Zároveň ale také mohou informace snadno sdílet prostřednictvím různých blogů, komentářů, či na sociálních sítích. Díky takovému množství informací sdílených online, lze texty analyzovat a hodnotit. Jednou z metod textové analýzy je analýza sentimentu.

Analýza sentimentu je metoda, která zkoumá lidské pocity, postoje, názory a emoce. Nezaměřuje se tedy na fakta, ale na subjektivní názory. Spotřebitelé sdílejí každou minutu svá hodnocení a to nejen produktů a služeb, ale i podniků a organizací. Mohou tak snadno a rychle informovat a ovlivňovat ostatní. Porozumění vnímání zákazníka může být příležitostí pro obchodníky, jež dokáží díky recenzím reagovat na zákaznickovy preference. Analýza sentimentu se však netýká jen komerčních společností. Využívá se také ve finanční oblasti, například na akciových trzích, čemuž je věnována tato diplomová práce.

Tato práce má za cíl zjistit vliv emocí v novinových člancích na hodnotu výnosu abnormální akcie. Tzn., zda emoce vyvolaná přečtením novinového článku ovlivní investora v nákupu cenných papírů.

V první části práce se nachází teoretické shrnutí problematiky, druhá část je praktická. Práce je rozdělena na čtyři hlavní kapitoly. V první kapitole je shrnuta tematika akciových trhů, včetně vysvětlení základních přístupů k predikci akciových výnosů – fundamentální, technické a psychologické analýzy.

Druhá kapitola se zabývá samotnou analýzou sentimentu. Nejdříve obecně popisuje její význam a definuje ji. Následně vysvětluje pojem názor, jeho druhy a prvky. Pokračuje dvěma hlavními přístupy analýzy emocí v textových dokumentech – slovníkovou metodou a metodou strojového učení. Všeobecná metodologie analýzy je využita v praktické části této práce. Na závěr kapitoly jsou zmíněny problémy a nedostatky analýzy sentimentu.

Třetí kapitola obsahuje sběr a zpracování dat, u kterých byl hodnocen sentiment. Celkem je zkoumáno deset společností, jejichž akcie jsou obchodovány na americké burze. Novinové články, týkající se těchto společností byly vyhledány pro období tří let 2016 až 2018. Takto stažená data ve formě textových souborů bylo potřeba očistit a připravit pro vstup

do programu STATISTICA Cz, ve kterém byly zjišťovány hodnoty sentimentu pomocí tří slovníků sentimentu obsahujících seznam pozitivních a negativních slov. Celkem byly použity slovníky od třech autorů - Bing Liu, Elaine Henry a dvojice Loughran a McDonald.

Výsledné hodnoty sentimentu jsou ve čtvrté kapitole použity k predikci abnormálních akciových výnosů pomocí regresní analýzy.

Získané poznatky a dosažení stanovených cílů je zhodnoceno v závěru práce.

1 AKCIOVÉ TRHY

Tato kapitola obsahuje teoretické shrnutí základních pojmů týkajících se akciových trhů. Nejdříve jsou popsány akcie a jejich druhy, následují přístupy k predikci akciových výnosů a na závěr jsou zmíněny burzy cenných papírů společně s vysvětlením pojmu abnormální akciový výnos.

1.1 Akcie

Význam podnikových akcií spočívá převážně v tom, že si akciové společnosti (dále jen a.s.) prostřednictvím jejich emitace a následným prodejem na primárním trhu opatřují potřebné peněžní prostředky, které nemusí v budoucnu vracet. Také přitahují zájem mnoha potenciálních investorů [25].

Akcie jsou cenné papíry (dále jen CP), díky nimž se jejich držitel stává akcionářem v dané akciové společnosti, čímž získává právo na podíl na zisku společnosti [9,10]. Ziskovost však nelze zaručit, je možné ji pouze předvídat. Proto si musí každý investor uvědomit, že v případě, že se podnik stane neziskový, může jeho příjem ve formě dividend klesnout nebo i zcela ustát. Navíc se může stát, že ačkoliv je podnik ziskový, jeho management se rozhodne zisky zadržet za účelem zajištění dostatečných fondů pro investice společnosti, z čehož by však v dlouhodobém horizontu měli mít akcionáři prospěch. Zadržení dividend tak bude spojeno s budoucími příjmy ze současné investice. V opačném případě (pokud by podnik zisky nezadržel a dividendy vyplatil) může být podnik nucen půjčit si cizí prostředky pro svoje fungování [17].

Dalším motivem k nákupu akcií bývá pro investory také snaha, dosáhnout kromě dividend i kapitálových výnosů, které vyplývají z předpokládaného růstu či poklesu jejich tržních cen. Akciové trhy tak neslouží jen k cílenému umístění dlouhodobých peněžních prostředků, ale i ke spekulacím [25].

Investoři tedy před nákupem akcií zvažují nejen pravděpodobnost výše dividend, ale i možný kapitálový výnos [17].

Ceny akcií se mění závisle na očekávání budoucích dividend. Pokud se předpokládá stálý a velký růst dividend, cena akcií může růst v reakci na zvýšenou poptávku. Jestliže se očekává snížení dividend, pak ceny akcií klesají. Při tvorbě očekávané ceny akcie investoři sledují převážně hlavní indikátory, a to z důvodu značného objemu informací. Nejčastějším

z nich je očekávaná volatilita cen akcií, která se přibližně posuzuje podle nejnižší a nejvyšší ceny akcie během roku [17].

Akcie jsou v tržní ekonomice nejrozšířenějším druhem cenných papírů. Je to dáno tím, že a.s. jsou [8]:

- nejčastější právní formou finančních a průmyslových společností;
- kapitálově nejsilnějšími společnostmi (mají nejvyšší základní kapitál);
- nejdůvěryhodnějšími společnostmi (nejpodrobněji právně určená společnost);
- společnostmi, které mohou získat kapitál od investorské veřejnosti.

1.1.1 Druhy akcií

Akcie jsou cenné papíry, díky nimž se jejich držitel stává spolupodílníkem v dané akciové společnosti. Pokud firma vydává první akcie, prodávají se za nominální cenu. Na burze je lze koupit za cenu tržní. Investor nákupem akcií vkládá do společnosti určitý kapitál a za to se pak může podílet na jejím zisku pravidelnou výplatou dividendy, ačkoliv je tato výplata proměnlivá. O její výši a výplatě hlasují každý rok akcionáři na valné hromadě. Výhodou pro akcionáře je i to, že neručí za závazky společnosti [6].

Akcie lze rozdělit podle [8]:

- **Emitace:**
 - Jednotlivé – tj. po jednom kusu akcií;
 - Hromadné – obsahují větší počet akcií, např. hromadná akcie na 100 kusů akcií. Akcionáři mají právo vyměnit hromadnou akcii za jednotlivé akcie.
- **Formy:**
 - Na jméno – akcie vždy nese jméno vlastníka (fyzické nebo právnické osoby), společnost vede seznam akcionářů, listinná akcie se převádí rubopisem (zapsání nového vlastníka na rubu akcie) a předáním;
 - Na majitele – akcie se převádí pouhým předáním, společnost neeviduje vlastníky akcií a nezná tak své akcionáře.

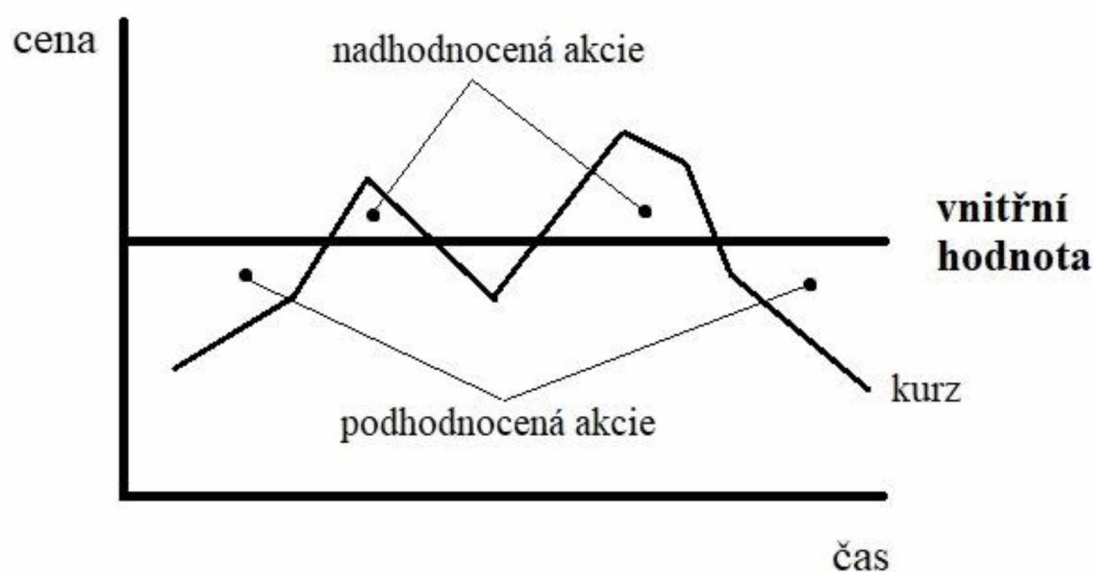
- **Druhu:**
 - Prioritní – mají přednostní právo na dividendu nebo likvidační zůstatek; souhrn prioritních akcií nesmí překročit 50 % základního kapitálu;
 - Kmenové – s nimiž nejsou spojena žádná zvláštní práva.
- **Podoby:**
 - Listinné – vytištěné akcie (papírové);
 - Zaknihované – v elektronické podobě vedené v zákonem stanovené evidenci (u nás v Centrálním depozitáři cenných papírů).

Při úpisu akcií vydává akciová společnost na upsané a dosud nesplacené akcie zatímní list. Vlastník zatímního listu má stejná práva jako akcionář. Po splacení všech upsaných akcií podnik vymění zatímní list za akcie [18].

Každá akcie může mít různou cenu a hodnotu. Jde o [8]:

- **Jmenovitou hodnotu** – hodnota uváděná na akcii (nominální).
- **Kurz** – tržní cena vzniklá na regulovaném trhu působením anonymní poptávky a nabídky po akcii. Kurz se odchyluje od jmenovité hodnoty, a to směrem nahoru i dolů. Kurz závisí na mnoha faktorech, z nichž nejdůležitější jsou:
 - výše dividendy (vyplacený podíl na zisku společnosti připadající na jednu akcii),
 - výše zisku (možnost financovat rozvoj společnosti vlastními zdroji),
 - rozvoj a růst společnosti (tržby, podíl na trhu, počet pracovníků).
- **Emisní kurz** – cena, za níž společnost vydává akcie (při úpisu). Emisní kurz nesmí být nižší než jmenovitá hodnota akcie, může však být vyšší. Rozdíl mezi emisním kurzem a jmenovitou hodnotou je emisní ážio a představuje zisk společnosti.
- **Smluvní cena**- cena sjednaná přímo mezi prodávajícím a kupujícím. Smluvní cena může být jiná, než je jmenovitá hodnota nebo kurz akcie.
- **Vnitřní hodnota** – skutečná hodnota pro akcionáře odpovídající ekonomické situaci a finančnímu zdraví společnosti. Určení vnitřní hodnoty slouží k investičnímu rozhodování. Je-li vnitřní hodnota vyšší než kurz, je akcie podhodnocena, je příliš levná a měla by se koupit. Je-li vnitřní hodnota nižší než

kurz, je akcie nadhodnocena, je příliš drahá a měla by se prodat. Vnitřní hodnota je graficky znázorněna na obrázku č. 1.



Obrázek 1 - Vnitřní hodnota akcie

Vlastní zpracování dle [8]

1.1.2 Burza cenných papírů

Akciový trh, známý též jako burza cenných papírů, je formálním trhem pro obchodování s CP společností. Akciový trh existuje k tomu, aby umožnil investorům nakupovat a prodávat stávající akcie, a aby společnosti mohly získávat peníze emitováním nových akcií. Burza je akciovou společností a platí pro ni ustanovení obchodního zákoníku [2,26].

Na burze se obchoduje prostřednictvím prostředníků a to z řad členů bank či makléřských firem. Obchodování může probíhat dvěma způsoby – prezenčně nebo elektronicky. Prezenční forma se vyznačuje obchodním parketem, kde zástupci členských institucí nahlas s burzovním specialistou vyjednávají nákup a prodej CP. Pro elektronickou formu je potom typické uzavírání obchodů automaticky prostřednictvím počítačového systému, do nějž jsou vkládány nákupní a prodejní příkazy, které se podle určitého kritéria vypořádají [2,26].

Vedle burzy existuje i další typ trhu CP-mimoburzovní (over-the-counter, dále jen OTC). Zde se obchoduje s takovými CP, které nesplnily kritéria pro přípuštění k burzovnímu obchodování. OTC může být organizovaný nebo neorganizovaný. Organizovaný OTC představuje organizovaný obchod s akciemi, které nebyly přípuštěny k burzovnímu obchodování. V posledních letech se však organizované mimoburzovní trhy vyvinuly natolik,

že řada z nich dnes představuje konkurenční a plnohodnotné trhy cenných papírů. Nejznámějším mimoburzovním trhem je americký NASDAQ, kde se obchoduje především s akciami technologicky orientovaných společností jako Apple, Microsoft, nebo Intel. V Česku je to pak RM-SYSTÉM [2].

Pojmem mimoburzovní trh se proto rozumí spíše neorganizovaný OTC, zahrnující CP, které nesplnily nároky pro vstup na některý z organizovaných trhů [2].

1.1.3 Burzovní index

Burzovní index je ukazatel, který popisuje dění na určitém trhu jako celek. Slouží ke sledování vývoje trhu v čase a jeho vývojových tendencí. Vypočítává se z cen vybraných podkladových aktiv, specifických pro daný trh. Zpravidla se vypočítává na bázi váženého průměru. V České republice se sleduje index PX (oficiální index Burzy cenných papírů Praha) a index RM (pro RM-SYSTÉM, českou burzu cenných papírů, a.s.) [26].

Na amerických akciových trzích se vyskytují tři nejsledovanější indexy - NASDAQ Composite, Dow Jones Average a S&P 500 [2].

1.2 Abnormální akciový výnos

Abnormální akciový výnos je míra výkonnosti, která je užitečná pro investory, jako nástroj pro oceňování a pro porovnávání výnosů s výkonností trhu. Na akciových trzích jsou abnormální akciové výnosy chápány jako rozdíly mezi výnosností jedné akcie nebo portfolia akcií a očekávaným výnosem trhu za stanovené časové období, přičemž jako měřítko k určení očekávaného výnosu se většinou využívá významný burzovní index, např. S&P 500 [1].

Abnormální návratnost investice se vypočítá následovně [11]:

$$AR_{it} = R_{it} - E(R_{it}) \quad (1)$$

kde AR_{it} je abnormální výnos firmy i za den t , R_{it} je výnos akcie firmy i za den t , a $E(R_{it})$ je očekávaný výnos trhu firmy i za den t . Abnormální výnosnost investice tak může být kladná nebo záporná.

Pokud například akcie společnosti vzrostly o 5 %, ale průměrný trh se zvýšil pouze o 3 %, pak byl abnormální výnos 2 % (5 % - 3 % = 2 %).

1.3 Současné přístupy k predikci abnormálních akciových výnosů

Oceňováním akciových výnosů se zabývají různé metody. Nejznámější z nich jsou fundamentální, technická a psychologická analýza, jež jsou podrobněji popsány níže.

1.3.1 Fundamentální analýza

Fundamentální analýza je považována za nejsouhrnnější druh akciové analýzy, jež se v investiční praxi používá při přípravě zásadních nákladových rozhodnutí [25].

Souhrnný záběr fundamentální analýzy je velmi široký. Kontroluje, do jaké míry cena akcie odpovídá její vnitřní hodnotě. Snaží se zjistit, které akcie jsou nadhodnocené, a které podhodnocené, čili které akcie koupit nebo prodat. Užívá exaktní početní metody posuzování finančních ukazatelů podniků. Zabývá se produktivitou společnosti v minulosti a to na základě aktivity společnosti jako celku. Nevylučuje ani použití formálních matematických modelů [18].

Základem pro pochopení podstaty fundamentální analýzy je správné chápání výrazu „vnitřní hodnota akcie“. Lze ji definovat jako osobní názor libovolného účastníka akciového trhu na to, jaký by měl být tzv. „spravedlivý kurz“, jehož hodnotu lze v ultrakrátkém období považovat za neměnnou, a je tedy možné ji porovnat s proměnným současným akciovým kurzem [25].

Z hlediska charakteru zkoumaných faktorů lze fundamentální akciovou analýzu rozdělit na tři typy [25]:

- 1) **Globální/ makroekonomická akciová analýza** – objasňuje ekonomiku jako celek a zabývá se zkoumáním vztahů mezi vývojem tzv. globálních, především makroekonomických agregátů a pohybem akciových kurzů.
- 2) **Odvětvová analýza**- úkolem těchto analýz je rozpoznávat a podrobně charakterizovat nejvýznamnější zvláštnosti jednotlivých odvětví a následně předpovídat perspektivy jejich budoucího vývoje.
- 3) **Analýza jednotlivých akciových společností a jejich akcií**- je analýza, která zkoumá vnitřní parametry konkrétních akciových společností v kontextu s vlastnostmi jimi vydávaných emisí akcií.

Jaké typy analýzy bude finanční analytik přesně provádět a jaké analytické metody k tomu použije, to ovšem záleží na mnoha skutečnostech. Zejména bude záviset na tom, z jakého důvodu analýzu provádí (jakou investiční strategii plánuje realizovat, v jakém časovém

horizontu a v jakém objemu). Dále bude záviset na tom, zda se bude jednat o tuzemskou investici či zahraniční, a v neposlední řadě také na tom, zda půjde o přímou investici s cílem získat v určité společnosti vlastnický podíl, nebo zda se bude jednat o portfoliovou investici usměrňovanou do některého oboru, či bude záviset pouze na faktorech rizikovosti, ziskovosti, popřípadě na likviditě jednotlivých vkladových (investičních) projektů atd. [25]

1) Globální/ Makroekonomická akciová analýza

Z makroekonomického hlediska ovlivňuje vývoj akciových trhů celá řada skutečností, z nichž jsou za nejvýznamnější považovány především [25]:

- a) reálný výstup ekonomiky,
- b) fiskální politika,
- c) peněžní nabídka,
- d) úrokové sazby,
- e) inflace,
- f) příliv a odliv zahraničního kapitálu,
- g) kvalita investičního prostředí.

a) Reálný výstup ekonomiky

Evoluce akciových trhů je vždy značně ovlivňována jak vývojem ekonomiky příslušného státu, tak i vývojem mezinárodní ekonomiky. Z této skutečnosti vyplývá, že také akciové kurzy reagují svými růsty či poklesy na probíhající ekonomické změny.

Z dlouhodobého hlediska však lze vyzorovat, že toto kolísání nastává kolem základního vývoje, který lze vyčíslit jako dlouhodobý, několikaprocentní roční ekonomický růst způsobený postupným růstem hospodářské úrovně dané ekonomiky [25].

b) Fiskální politika

Fiskální politikou rozumíme správu a vedení příjmů, výdajů a dluhu vlády určitého státu. Především daně z příjmů (jak právnických, tak i fyzických osob) mají významný vliv na atraktivnost akcií. Snižují dosažené (hrubé) zisky firem, což má negativní vliv na výši vyplácených dividend, a to omezuje možnosti následného rozvoje akciových společností [25].

c) Peněžní nabídka

Peněžní nabídka je běžně považována za jeden z nejdůležitějších faktorů ovlivňujících vývoj na akciových trzích. Podle finanční teorie rozlišujeme tři vysvětlení jejího působení [25]:

- První způsob vysvětlení – zakládá se na názoru, že zprostředkuje-li centrální banka v rámci své monetární politiky zvýšení peněžní nabídky při zachování konstantní poptávky po penězích, investoři tyto doplňkové peněžní prostředky investují částečně i na akciových trzích. Následně dochází ke vzestupu kurzů akcií. Tento přímý vliv peněz na akciové kurzy nazýváme „efekt likvidity“.
- Druhý způsob vysvětlení – zakládá se na předpokladu, že růst finanční nabídky nejdříve iniciuje investory nakupovat více dluhopisů, což způsobí zvýšení jejich tržních cen. Ten se ovšem ihned projeví v poklesu výnosnosti, čímž stoupne přitažlivost možných aktiv, kterými jsou v tomto případě akcie. Výsledkem tedy bude zvýšení poptávky po akciích a následující nárůst jejich tržních cen.
- Třetí způsob vysvětlení – zvýšení peněžní nabídky vyvolá snížení úrokových sazeb, což podnítlí investiční činnost firem. Tím se zvýší jak jejich předpokládané výnosy, tak velikosti v budoucnu vyplácených podílů na zisku, a tím dojde k růstu tržních cen akcií.

d) Úrokové sazby

Dalším důležitým faktorem akciových trhů jsou změny úrokových sazeb. Platí, že jejich růst vede k poklesu tržních cen akcií.

To lze vysvětlit různými způsoby [25]:

- První způsob vysvětlení tkví v již uvedeném názoru, že současné hodnoty akcií závisí na současných hodnotách budoucích peněžních příjmů z nich plynoucích. Navýšení sazby musí vést ke snížení současné hodnoty nastávajících akcionářů, což zapříčiní pokles jak „vnitřních hodnot“ akcií, tak i s nimi spřízněných akciových kurzů.
- Druhý způsob vysvětlení vychází z osvědčené zkušenosti, že zvýšení úrokových sazeb a tím i běžných výnosů plynoucích z dluhopisů odebírá z akciových trhů peněžní prostředky na koupi dluhopisů, které umožňují domct se vyšší

výnosnosti. A proto až do doby, než se vytvoří nová stabilita v ziskovosti na obou těchto trzích, budou kurzy klesat.

- Třetí způsob vysvětlení se zakládá na skutečnosti, že vyšší úrokové sazby navyšují výdaje firem na jejich externí dotování. Předpokládá se snížení očekávání budoucích zisků a pokles tržních cen.

e) Inflace

Výzkumy prokázaly, že při inflaci akcie většinou nejsou schopny dlouhodobě udržet svoji reálnou hodnotu, a to i přesto, že představují vlastnické právo k reálným aktivům a.s. V dnešní době existuje řada hypotéz, které se tuto určitou zvláštností snaží vysvětlit. Nejznámější z nich je „hypotéza daňového efektu“, která se zabývá odepisováním investičního kapitálu a oceňováním zásob firmy v době inflace [25].

f) Příliv, resp. odliv zahraničního kapitálu

Změny v pozornosti zahraničních investorů mají neobvykle silný vliv na vyvíjení akciových kurzů. Příliv financí ze zahraničí způsobuje růst na domácím akciovém trhu, odliv zahraničního kapitálu způsobuje naopak pokles. Navíc, když se vezme v úvahu, že akcie významných korporací se stále častěji obchodují současně na více mezinárodních burzách, způsobuje jejich vzestup (pokles) na jednom trhu následně vzestupy (poklesy) i na ostatních trzích, a to v celosvětovém rozsahu [25].

g) Kvalita investičního prostředí

Lze ji charakterizovat jako všeobecné podmínky, které vzbuzují nebo naopak nevzbuzují důvěru případných investorů k realizaci jimi zamýšlených investic. Investoři věnují pozornost především transparentnosti, bezpečnosti a stabilitě finančního trhu (což pochopitelně souvisí s celkovou vyspělostí dané země).

Existuje pochopitelně mnoho faktorů podílejících se na kvalitě investičního prostředí, nicméně v centru pozornosti investorů bývají především [25]:

- schodky a nerovnováha,
- korupce, hospodářská kriminalita a právní systém,
- cenové regulace a černý trh,
- ekonomické a politické šoky.

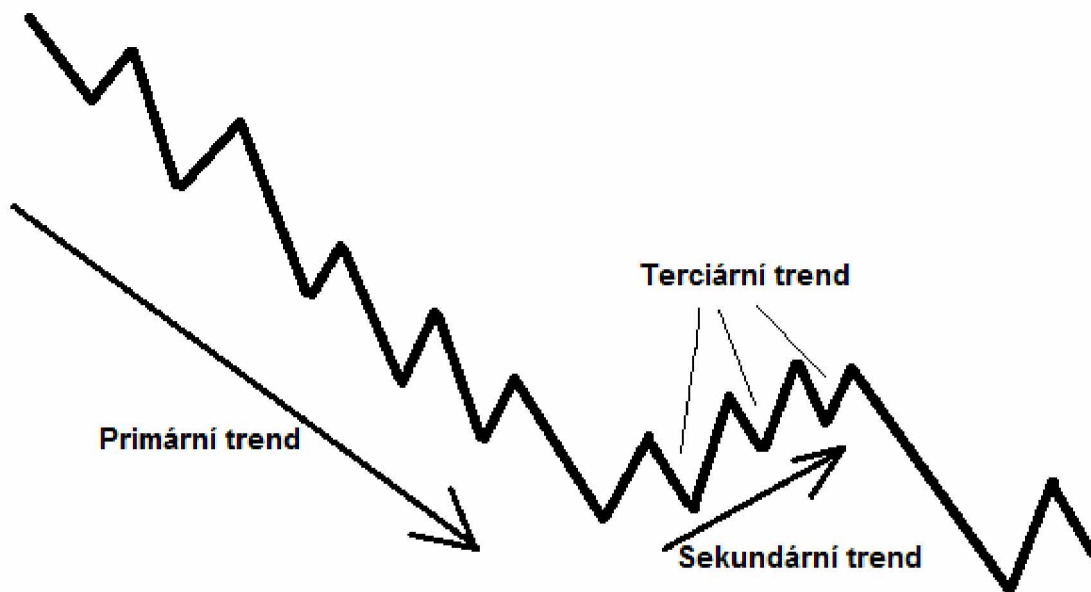
1.3.2 Technická analýza

Technická analýza se zaměřuje jen na dění na finančním trhu, tj. na minulé a současné informace o cenách akcií a objemu transakcí. Usiluje o předpovídání vývoje cen jednotlivých akcií i celkového vývoje trhu. Hlavním cílem je posoudit s předstihem úpravy v trendech cen akcií. Hledá odpověď na otázku, kdy prodej nebo nákup akcií uskutečnit [17].

Základní východiska technické analýzy [17]:

- V tržních cenách akcií se odráží vše, co je může ovlivnit. Technická analýza zapisuje v grafické podobě vývoj cen akcií a z toho usuzuje následující trend.
- V trendech se vyvíjejí ceny, trendy mají určitou míru neměnnosti. Cena akcie se pohybuje v daném směru, ve kterém se pohyb oslabí a pak se pohybuje v jiném trendu.
- Vývojové etapy a uspořádání se opakují, tento princip vychází z lidské podstaty reagovat za stejných okolností podobně.

Technická analýza je schopnost předpovídat změny cenových trendů dříve, než k nim dojde. Zásadním trendem je tzv. **bázický trend**, který trvá přinejmenším několik let. Jeho vymezení umožňuje investorovi určit možnosti vhodných investic při dlouhodobých rozhodnutích. Na bázický trend navazují následující trendy. **Primární trend** (dlouhodobé „býčí“ a „medvědí“ trendy – *bull and bear trends*) trvá od jednoho roku do několika let. Je ovlivňován kratšími tzv. **sekundárními trendy** (*secondary trends*), s dobou trvání cca od několika týdnů až do několika měsíců, a **trendy terciálními** (*minor trends*), které mají dobu trvání od několika dní až po několik týdnů. Tuto teorii poprvé objasnil na konci 19. století editor *The Wall Street Journal* Charles Dow [18].



Obrázek 2 - Primární, sekundární a terciální trend - teorie Charles Dow

Zdroj: [19]

Podle toho jaký „primární“ trend v dané době převládá, se akciový trh označuje následujícími způsoby [6]:

- Rostoucí (býčí) trh – vyznačuje se růstem daného burzovního koeficientu, a tím i růstem většiny kurzů akcií,
- Klesající (medvědí) trh – vyznačuje se svým poklesem, to znamená jak poklesem daného burzovního indexu, tak i poklesem kurzů většiny akcií,
- Postranní trh – označuje se také jako trh bez trendu, tzv. „*sideway market*“, vyznačuje se horizontálním vývojem.

Začátkem cyklu označujeme tzv. *fázi akumulace*, po jejím skončení trh přechází do růstové fáze, během níž vykazuje přechodné pauzy či chvilkové poklesy označované jako *korekce*. Na konci vzestupné fáze pak trhy vrcholí, to se projevuje opět časově omezeným vedlejším trendem označovaným jako *distribuce*. Poté trh přechází do sestupné (medvědí) fáze, která bývá prudší, a tedy i kratší než fáze vzestupná, a přitom může vykazovat dočasné korekce, a končí opakovaným nástupem fáze akumulace [25].

Dowova teorie se však dočkala i mnoha kritických připomínek, nejdůležitější z nich jsou [6]:

- signály k nákupům a prodejům jsou indikovány příliš pozdě,
- často se stává, že signály nejsou jednoznačné,
- tato teorie nedefinuje ani dobu trvání, ani objem jednotlivých (primárních, sekundárních a terciálních) trendů,
- jedná se o teorii, která se zabývá akciovým trhem jako celkem, a v tom případě neumožňuje dobře posuzovat jednotlivé akciové trhy.

Dowovu teorii lze zhodnotit jako první souhrnnou teorii zabývající se technickými postupy, přinesla základní poznatky potřebné pro vypracování prvních použitelných metod.

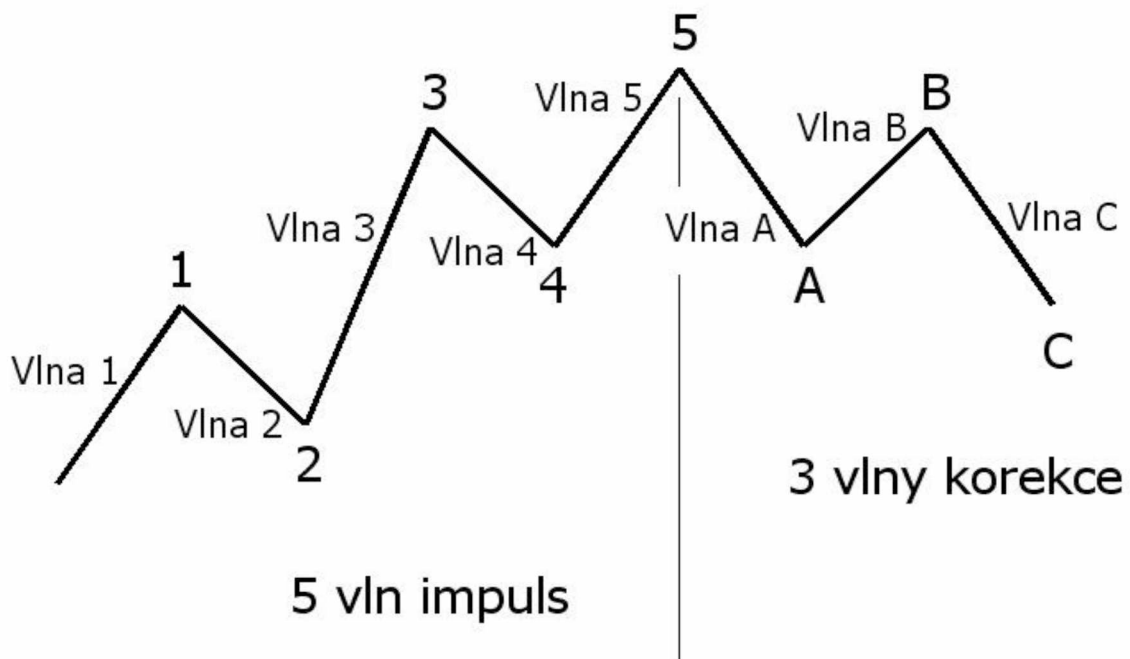
Současné chápání Dowovy teorie je ve srovnání s jeho původní teorií už trochu odlišné, v dnešní době se považuje za možné používat technickou analýzu i při velice krátkodobých investičních strategiích, takže se využívají i sekundární, případně i terciální trendy [25].

Teorie Elliottových vln

Tato teorie vychází z domněnky, že v přírodních jevech existují určitá období, která lze nalézt i v ekonomice (např. střídání období konjunktury s obdobím recese). Z toho je možné vyvodit, že změny akciových burzovních indexů vycházejí z přirozenosti lidského chování a pohybují se v rozlišitelných cenových rámcích – v tomto případě ve vlnách, které mají vnitřní strukturu a je možné je klasifikovat na různých posloupných úrovních. Snaží se jednak definovat pravidla vztahu mezi posloupnostmi vln, nejdříve graficky, resp. numericky určovat vzdálenosti mezi nimi a poté odhadovat budoucí změny akciového trhu [9].

Elliottova teorie pracuje s vlnovými formacemi, které se neustále opakují. Jsou složeny z tzv. „základního fraktálu“, neboli ze vzestupného a sestupného pohybu, které popisují charakter akciového trhu jako celku. Je postavena na tvrzení, že trh vzrůstá v „impulzní vlně“, která se skládá z pěti pod-vln, označovaných číselně 1, 2, 3, 4, 5, za ní následuje „vlna korekční“, skládající se ze tří pod-vln označovaných písmeny a, b, c. Vlny ve směru hlavního trendu jsou označovány jako „impulzní vlny“, a vlny, které působí v protisměru, neboli proti hlavnímu trendu, jako „vlny korekční“ [25].

Základní fraktál Elliottových vln je znázorněn na obrázku č. 3.



Obrázek 3 - Základní rámec Elliottových vln

Zdroj: [20]

Při analyzování libovolné Elliottovy vlny je vždy nejprve potřeba zjistit, jaká je její nejbližší nadřazená vlna [21]**Chyba! Nenalezen zdroj odkazů.**:

- Jestliže je její nadřazená vlna impulzní, tak patří do skupiny jejich pěti podřazených vln, z nichž vlny 1, 3 a 5 představují impulzní vlny, které působí ve směru hlavního trendu a vlny 2 a 4 představují korekční vlny působící opačně.
- Jestliže je její nadřazená vlna korekční, tak je jednou ze tří jejích pod-vln, u nichž vlna (a) a vlna (c) představují impulzní vlny působící ve směru nadřazené vlny a prostřední vlna (b) představuje opačně působící pod-vlnu korekční.

1.3.3 Psychologická analýza

Zakládá se na předpokladu, že budoucí vývoj kurzů akcií závisí na impulzech, které ovlivňují chování davu, jenž je jimi veden buď k nákupům, nebo k prodejům.

Důležitost kolektivní psychologie dokládá známý názor Keynese, který již ve své době tvrdil, že „nemá žádného smyslu zaplatit 25 peněžních jednotek za investici, o níž podle jejího perspektivního výnosu důvodně soudíte, že má sice hodnotu 30, avšak předpokládáte, že za tři měsíce ji trh ocení pouze na 20“ [25].

Z jeho názoru vyplývá, že je třeba, aby investor věnoval velkou pozornost předvídání následných změn v prostředí psychologie trhu a vhodným způsobem využíval svých odhadů [25].

Nejvýznamnější teoretické koncepce zabývající se vlivem masové psychologie na kurzy akcií [8]:

- a) Keynesova spekulativní rovnovážná hypotéza,
- b) teorie spekulativních bublin,
- c) Kostolanyho burzovní psychologie,
- d) Drasnarova koncepce psychologické analýzy.

a) Keynesova spekulativní rovnovážná hypotéza

Je nejstarší teorií psychologické analýzy, jejíž základy položil Keynes. Jako první v historii považuje spekulativní chování investičního publika za jeden z nejvýznamnějších faktorů působících na kurzy akcií [25].

b) Teorie spekulativních bublin

Někdy se stává, že kurzy na akciových trzích po určitou dobu rostou (příp. klesají) bez toho, aby bylo možné pro jejich vývoj nalézt rozumné fundamentální vysvětlení. Po určité době se však tento růst (příp. pokles) neočekávaně zastaví a kurzy zamíří velice prudce dolů. Stav, kdy se akciové kurzy dočasně, bez jakéhokoli racionálního vysvětlení velmi výrazně odchyľují od vnitřních hodnot stanovených fundamentálními analytiky, se nazývají „spekulativní bubliny“ a jejich objevení je vysvětlováno kolektivní psychologií [25].

c) Kostolanyho burzovní psychologie

Vychází z předpokladu, že v krátkém období (období do 1 roku) jsou kurzy akcií ovlivňovány především psychologickými reakcemi burzovního publika na různé události. Základem Kostolanyho burzovní psychologie je v první řadě rozdělení účastníků burzovního obchodování podle charakteristických znaků jejich chování a poté, na základě toků akcií a peněz mezi nimi, dále definování zásad pro odhadování budoucích kurzových změn [25].

Kostolanyho psychologie rozlišuje dva typy účastníků burzovního obchodování [25]:

- **Hráči** - investoři, kteří chtějí velmi rychle dosáhnout kurzových zisků, proto pouze reagují na nové informace a události. Chovají se emocionálně a jdou vždy s proudem, tzn. kupují, když všichni kupují, a prodávají, když ostatní prodávají. Hráči se orientují na malé zisky a jednají krátkodobě.
- **Spekulanti** – realizují dlouhodobější transakce, na rozdíl od hráčů, a počítají s rozsáhlejšími kurzovými pohyby. Mají vlastní myšlenky, prognózy, představy a nechovají se emocionálně, většinou jdou proti proudu. Jejich odhady následných pohybů kurzů akcií jsou podloženy konkrétními argumenty. Podle zkušeností jsou spekulanti na burzách mnohem úspěšnější než hráči.

d) Drasnarova koncepce psychologické analýzy

Drasnarova koncepce objasňuje proces zhodnocování a znehodnocování kurzů akcií jako následek působení dvou opačných vlastností, kterými podle Drasnara disponuje každý člověk. Jedná se o chamtivost a strach. Dle toho, která vlastnost převládá, dochází na akciových trzích k růstu či poklesu [25].

Chamtivost vzniká z běžného přání člověka něco vlastnit. Tato touha nikdy nemizí, ale s přibývajícím majetkem (bohatstvím) se spíše násobí. Snaha dosahovat vysokých zisků z investic, která je podporována chamtivostí, povzbuzuje poptávku po akciích, to vede následně k růstu jejich kurzů. Tento znehodnocovaný trend trvá až do momentu, kdy je chamtivost investorů nahrazena strachem ze ztráty získaného bohatství. Podobně, jako je chamtivost investorů neustále střídána strachem, dochází na trzích k neustálému střídání vzestupných a sestupných trendů. Ty se střídají, v závislosti na tom, jak se mění pocity davu [25].

2 ANALÝZA SENTIMENTU

Rozvoj analýzy sentimentu úzce souvisí s rozvojem internetu, obzvláště s masivním rozšířením sociálních sítí jako jsou například Facebook, Twitter, nebo Instagram. Lidé napříč národnostmi, pohlavím, věkem, rasou a sociální třídou používají internet k šíření svých zkušeností a dojmů ke každému aspektu jejich života. Ať už jde o posílání emailů, blogování, komentování na webových stránkách či sdílení na sociálních sítích, miliony lidí vyjadřují tímto způsobem svůj názor každý den.

Podle výzkumu společnosti Domo [12] lze zjistit, že v roce 2018 bylo každou minutu posláno přes 159 362 760 emailů, sdíleno 473 400 tweetů, či přidáno 49 380 fotografií na Instagram.

Tyto příspěvky kolují skrz online komunity a zákazníci tak mohou snadno a rychle informovat a ovlivňovat ostatní. Každý den tak vznikají nespočetné data informací, která umožňují marketérům nahlédnout hluboko do tendencí chování zákazníka a představují příležitost pro poznání emocí, pocitů a vnímání zákazníka v současnosti [20].

Velkou výhodou analýzy sentimentu je, kromě snadno získatelného obrovského množství dat, také menší finanční nákladnost v porovnání s ostatními marketingovými technikami, jako jsou rozhovory, dotazníky či telefonáty [11].

Analýzu sentimentu však nevyužívají jen komerční společnosti. Monitorováním sentimentu na sociálních sítích se snaží zlepšit svůj volební výsledek i politické strany. Podle mínění a očekávání svých voličů tak mohou přizpůsobit svoji volební kampaň a dosáhnout tak daleko většího úspěchu [23].

Analýza sentimentu se také využívá na akciových trzích. Provedením analýzy sentimentu ve zprávách o společnostech je možné odhadnout vývoj výnosnosti jejich akcií. Obecně, pokud je sentiment pozitivní, dochází k držení nebo nákupu akcií. Jestliže je sentiment negativní, dochází k jejich prodeji. Právě touto částí analýzy sentimentu se bude zabývat tato diplomová práce.

2.1 Definice

Definici analýzy sentimentu lze nalézt například v knize Liu Binga *Sentiment Analysis and Opinion Mining* [1]: „*Analýza sentimentu, často nazývaná jako dolování názorů (opinion mining), je vědní obor, který analyzuje lidské názory, pocity, hodnocení, postoje a emoce vůči*

subjektům, jako jsou výrobky, služby, organizace, jednotlivci, problémy, události, témata a jejich atributy.“

Gama a Rambocas vysvětlují ve své knize Marketing research: The role of sentiment analysis [2][20] analýzu sentimentu takto: „*Analýza sentimentu je technika dolování dat, která používá počítačové zpracování přirozeného jazyka, analýzu textu a matematickou lingvistiku, aby identifikovala a získala názory z textových dat.*“.

Analýza sentimentu je tedy jednou z metod analýzy textu, která slouží ke zjištění názorů a sentimentu lidí na různé otázky. Může jít nejen o recenze produktů a služeb, ale i hodnocení podniků, organizací, politických stran a dalších [23].

2.2 Názor

Názory jsou klíčové pro téměř všechny lidské činnosti a jsou klíčovými činiteli našeho chování. Naše přesvědčení a vnímání reality jsou do značné míry podmíněné tím, jak ostatní vidí a hodnotí svět. Proto často hledáme názory druhých, když se potřebujeme rozhodnout. To neplatí pouze pro jednotlivce, ale i pro organizace. Hlavním předmětem studia analýzy sentimentu jsou tedy názory a související pojmy, jako pocity, postoje, emoce a hodnocení[1].

2.2.1 Druhy názorů

Názory lze rozdělit podle dvou hledisek na běžné a srovnávací nebo explicitní a implicitní.

Běžný názor – běžný názor je často v literatuře označován jen jako názor a má dva podtypy [19]:

- **Přímý** – přímý názor vyjadřuje názor přímo na předmět sledování, např.: „Kvalita obrazu je skvělá“,
- **Nepřímý** – nepřímý názor je vyjádřen nepřímo entitou nebo aspektem entity na základě jejich účinků na některé jiné entity. Tento podtyp se často vyskytuje v lékařské oblasti. Například věta „Po použití těchto léků, se bolest kloubů zhoršila.“ popisuje nežádoucí účinek léku na klouby, které nepřímo dávají negativní stanovisko nebo sentiment k léku.

Velká část současného výzkumu se zaměřuje na přímé názory, jelikož je snadnější s nimi manipulovat. S nepřímými názory je často těžší se vypořádat. Například v oblasti léků je třeba vědět, jaký je žádoucí a nežádoucí stav před nebo po použití léčiva. Například věta "Vzhledem k tomu, že moje klouby byly bolestivé, můj lékař mi předepsal tento lék"

nevyjadřuje sentiment nebo názor na lék, protože "bolestivé klouby" (což je negativní) byly bolestivé ještě před použitím léku.

Srovnávací názor - srovnávací názor vyjadřuje vztah podobností nebo rozdílů mezi dvěma nebo více subjekty a preferencí nositele stanoviska na základě některých sdílených aspektů. Například věty „iPhone je lepší než Huawei.“ a „iPhone je nejlepší mobilní telefon.“ vyjadřují dva srovnávací názory.

Explicitní názor – explicitní názor vyjadřuje subjektivní stanovisko, které podává běžný nebo srovnávací názor. Například: „iPhone je skvělý.“, nebo „iPhone je lepší než Huawei.“.

Implicitní názor - implicitní názor je objektivní prohlášení, které předpokládá běžné nebo srovnávací stanovisko. Takové objektivní vyjádření obvykle vyjadřuje žádoucí nebo nežádoucí skutečnost, např. "Před týdnem jsem si koupil matraci a vytvořila se v ní proleženina" a "Životnost baterií telefonů Nokia je delší než u telefonů Samsung".

Explicitní názory jsou snazší odhalit a klasifikovat než implikované názory. Velká část současného výzkumu se tedy zaměřuje právě na explicitní názory.

2.2.2 Prvky názoru

Jak již bylo zmíněno výše, názor může obsahovat pozitivní či negativní sentiment. Sentiment však není jediným prvkem, který lze v názoru vyčlenit. Na příkladu recenze fotoaparátu Canon, použité z knihy [19], budou vysvětleny i ostatní komponenty názoru.

Autor: John Smith

Datum: 10. září 2011

„(1) Koupil jsem fotoaparát Canon G12 před šesti měsíci. (2) Prostě ho zbožňuji. (3) Kvalita obrazu je úžasná. (4) Výdrž baterie je také velká. (5) Avšak moje manželka si myslí, že je na ni moc těžký.“

Z tohoto hodnocení lze vyčíst několik důležitých bodů. Tato recenze obsahuje několik názorů, pozitivních i negativních, o fotoaparátu Canon G12. Věta (2) vyjadřuje pozitivní názor na fotoaparát Canon jako celek. Věta (3) vyjadřuje pozitivní názor na kvalitu obrazu tohoto fotoaparátu. Věta (4) vyjadřuje pozitivní názor na životnost baterie. Věta (5) obsahuje negativní názor na hmotnost kamery. Z těchto vět lze vyčíst, že názor se skládá ze dvou klíčových složek: cíl (**g**) a sentiment cíle (**s**), tj. (**g, s**), kde **g** může být jakákoli entita nebo aspekt entity, o které byl vyjádřen názor, a **s** je pozitivní, negativní nebo neutrální sentiment, nebo numerické ratingové skóre vyjadřující intenzitu sentimentu (např. 1-5 hvězdiček). Pojmy

pozitivní, negativní a neutrální vyjadřují orientaci nebo polaritu sentimentu. Cílem názoru je například ve větě (2) Canon G12 a ve větě (3) kvalita obrazu Canonu G12.

Tato recenze také obsahuje názory dvou osob. Autorem názoru ve větách (2), (3) a (4) je John Smith, autorem názoru ve větě (5) je však jeho žena.

Dalším důležitým prvkem je datum vytvoření recenze. V průběhu času totiž dochází k vývoji názoru. Dnes by měl autor recenze jistě jiný názor na fotoaparát Canon G12, než v době vzniku recenze, tedy 10. září 2011.

Z výše vyplývajícího tedy může být zřejmé, že názor se skládá ze čtyř hlavních složek, tzv. názorové čtveřice (**g,s,h,t**), kde **g** je cíl názoru, **s** vyjadřuje sentiment, **h** (holder) je držitel, neboli autor názoru a nakonec **t** (time), což vyjadřuje čas, ve kterém byl názor vyjádřen.

Tato definice, i když poměrně stručná, nemusí být snadno použitelná v praxi, zejména v oblasti online recenzí produktů, služeb a značek, protože úplný popis cíle může být složitější a nemusí se ani objevit ve stejné větě. Například ve větě (3), je cíl názoru vlastně „kvalita obrazu Canon G12“, ale věta zmínila jen „kvalitu obrazu“. V tomto případě není cílem mínění pouze „kvalita obrazu“, protože bez vědomí, že věta vyhodnocuje kvalitu obrazu kamery Canon G12, je samotný názor ve větě (3) málo užitečný. V praxi může být cíl často rozložen a popsán strukturovaným způsobem s více úrovněmi, což značně usnadňuje jak získávání názorů, tak i pozdější využití výsledků daných názorů. Například "kvalita obrazu Canon G12" může být rozložena do entity a atributu entity a reprezentována jako pár: (Cannon-G12, kvalita obrazu).

Entita **e** tedy může být produkt, služba, téma, předmět, osoba, organizace nebo událost. Je popsána dvojicí, **e: (T, W)**, kde **T** je hierarchie částí, dílčích částí atd. A **W** je množina atributů entity **e**. Každá část nebo dílčí část má také svůj vlastní soubor atributů. Entitou může být třeba konkrétní model fotoaparátu, např. Canon G12. Má sadu atributů, např. kvalitu obrazu, velikost a hmotnost, a sadu částí, např. čočky, hledáčky a baterie. Baterie má však také vlastní sadu atributů, např. životnost baterie a hmotnost baterie.

Pokud se tedy zváží rozdělení cíle názoru na entitu a atribut entity, vznikne nová definice názorové pětičky, kdy lze názor vyjádřit pěti složkami (**e_i, a_{ij}, s_{ijkl}, h_k, t_l**), kde **e_i** je název entity, **a_{ij}** je aspekt entity **e_i**, **s_{ijkl}** je sentiment na aspekt **a_{ij}** entity **e_i**, **h_k** je držitelem stanoviska a **t_l** je doba, kdy je vyjádření vyjádřeno držitelem **h_k**. Sentiment **s_{ijkl}** je pozitivní, negativní, nebo neutrální, nebo vyjádřený různými úrovněmi intenzity na bodovací škále (např. 1-5 hvězd).

V této definici jsou záměrně použity indexy pro zdůraznění, že pět informací v pětici musí odpovídat. To znamená, že sentiment s_{ijk} musí vydat držitel stanoviska h_k o aspektu a_{ij} entity e_i v čase t_i . Jakýkoliv nesoulad by byl chybný.

Podstatných je všech pět složek. Kdyby chyběla jen jediná z nich, vznikl by problém. Pokud například není známa časová složka, nelze analyzovat názory na subjekt podle času, což je v praxi často velmi důležité, protože názor vyjádřený před dvěma lety není stejný jako názor vyjádřený dnes. Problém vzniká i při absenci držitele sentimentu. Například, ve větě “Starostu milovali všichni lidé ve městě, ale zároveň byl kritizován vládou.” jsou dva držitelé sentimentu - “lidé ve městě” a “vláda”, kteří drží opačný názor na skutečnost.

Tato definice poskytuje rámec pro transformaci nestrukturovaného textu na strukturovaná data. Názorová pětice výše je v podstatě databázové schéma, na jehož základě lze extrahované názory vkládat do databázové tabulky, pomocí které je možné provádět celou řadu kvalitativních, kvantitativních a trendových analýz názorů.

2.3 Úrovně analýzy sentimentu

Analýzu sentimentu lze provádět na třech úrovních[19]:

Úroveň dokumentu - úkolem na této úrovni je klasifikovat, zda celý dokument vyjadřuje pozitivní nebo negativní sentiment. Například při přezkoumání hodnocení produktu systém určí, zda průzkum vyjadřuje celkový pozitivní nebo negativní názor na produkt. Tato úroveň analýzy předpokládá, že každý dokument vyjadřuje názory na jeden subjekt (např. jeden výrobek). Není tedy použitelný pro dokumenty, které hodnotí nebo porovnávají více entit.

Úroveň věty – úkol na této úrovni analyzuje jednotlivé věty a určuje, zda každá věta vyjádřila pozitivní, negativní nebo neutrální názor. Neutrální obvykle znamená žádný názor. Tato úroveň analýzy úzce souvisí s klasifikací subjektivity, která rozlišuje věty objektivní (vyjadřují věcné informace z vět) a subjektivní (vyjadřují subjektivní názory).

Úroveň entit a aspektů - analýzy na úrovni dokumentu i na úrovni věty nezjistí, co se lidem líbí a co se jim nelíbí. Úroveň aspektu provádí jemnější analýzu. Namísto pohledu na jazykové konstrukce (dokumenty, odstavce, věty nebo fráze) se úroveň aspektu přímo zabývá samotným názorem. Je založena na myšlence, že názor se skládá ze sentimentu (pozitivního či negativního) a cíle. Názor má bez určení jeho cíle omezené využití.

Například, ačkoliv má věta: „Ačkoli obsluha není tak skvělá, stále mám rád tuto restauraci.“ jasně pozitivní tón, nemůžeme říci, že tato věta je zcela pozitivní. Ve skutečnosti je věta pozitivní o restauraci, ale negativní o jeho obsluze. V mnoha aplikacích jsou cíle

mínění popsány entitami a jejich různými aspekty, a proto cílem této úrovně analýzy je odhalit city o subjektech anebo o jejich aspektech. Například, věta "Kvalita hovoru telefonu iPhone je dobrá, ale životnost jeho baterie je krátká." vyhodnocuje dva aspekty: kvalitu hovoru a životnost baterie telefonu iPhone. Sentiment na kvalitu volání iPhone je pozitivní, ale sentiment na životnost baterie je negativní. Kvalita hovoru a životnost baterie iPhone jsou stanoveny jako cíle. Na základě této analýzy lze vytvořit strukturované shrnutí názorů na subjekty a jejich aspekty, které promění nestrukturovaný text na strukturovaná data, která mohou být použita pro všechny druhy kvalitativních a kvantitativních analýz.

2.4 Metody analýzy sentimentu

Analýza sentimentu může být prováděna několika metodami. V této kapitole budou vysvětleny dva základní směry – metoda strojového učení a slovníkové metody.

2.4.1 Strojové učení

Strojové učení, nazýváno také ML – machine learning, je spojeno s oblastí umělé inteligence a zaměřuje se na budování výpočetních modelů z minulých zkušeností a pozorování. Tato metoda zásadně podporuje využívání počítačového programování, aby se naučila a porozuměla základním datům a pak využila získané znalosti k předvídání nebo optimalizaci některých budoucích kritérií. Obecným cílem je vytvořit prediktivní funkci schopnou předpovědět cílový výsledek - y (závislá proměnná) pomocí předdefinovaných vstupních kritérií nebo atributů [23].

Když je tento cíl znám, tento typ učení se nazývá supervizní učení. Využití supervizního učení v analýze sentimentu vyžaduje tréninkový dokument textového obsahu nebo datového korpusu, který slouží jako přípravný dokument pro klasifikační učení. Tři základní techniky pro klasifikaci zahrnují: Naive Bayes, Support Vector Machines a Maximum-Entropy [20].

2.4.2 Slovníkové metody

Tyto metody vyhledávají v analyzovaném textu slova vyjadřující sentiment a podle nich vyhodnocují sentiment celého zkoumaného textu. Tato slova, nazývaná také jako názorová slova se běžně používají k vyjádření pozitivních či negativních pocitů. Například slova: „dobrý, skvělý, úžasný“ vyjadřují pozitivní sentiment a slova: „špatný, ubohý, hrozný“ vyjadřují naopak sentiment negativní. Názorová slova jsou převážně přídavnými jmény či příslovci, mohou však být použita i podstatná jména (např.: „odpad, hrůza, hloupost“)

či slovesa (např.: „milovat“ a „nenávidět“). Kromě jednotlivých slov se využívají i fráze a idiomy (např.: „Stojí to za houby.“, „Nemůžu bez toho žít.“)[23].

Tyto slova rozřazena podle polaritý jsou následně seřazena do dvou seznamů. Těmto seznamům se říká slovník sentimentu (sentiment lexicon).

Slovníky lze vytvářet manuálně nebo automatizovaně. Upřednostňují se spíše automatizované přístupy, neboť manuální přístup je zdoluhavý a náročný. V případě automatizovaného přístupu se začne s několika slovy, kterým je určen sentiment manuálně. K nim jsou vyhledávána synonyma a antonyma, kde synonymu je přiřazen stejný sentiment jako základnímu slovu a antonymu je přiřazen opačný sentiment. Tímto způsobem vznikne rozsáhlý seznam slov vyjadřujících sentiment, který umožňuje provedení analýzy sentiment [23].

Takovýchto slovníků existuje hned několik. Většina z nich vychází z původního Harvardského slovníku s označením Harvard General Inquirer IV-4. Mezi takovéto patří například slovníky Loughrana a McDonalda, Eline Henry, či Binga Lui.

Ačkoli jsou názorová slova a fráze pro analýzu sentimentu důležitá, jejich použití není zdaleka dostačující. Problém je mnohem složitější. Například při použití negace ve větě, či užití sarkasmu nebo ironie. Můžeme tedy říci, že lexikon slovní zásoby je nezbytný, ale ne zcela postačující. Tato problematika bude podrobněji popsána v kapitole 2.7 Problémy analýzy sentimentu.

2.5 Metodologie

Analýza sentimentu se provádí v pěti krocích [20]:

- sběr dat,
- příprava textu,
- detekce sentimentu,
- klasifikace sentimentu,
- prezentace výstupů.

2.5.1 Sběr dat

Analýza sentimentu využívá rozsáhlého, přes internet uživatelsky generovaného obsahu. Zdrojem dat jsou většinou dotazy uživatelů na veřejných fórech, jako jsou blogy, diskuzní fóra a přehledy s recenzemi produktů, jakož i na soukromých stránkách prostřednictvím

sociálních sítí, jako je Twitter a Facebook. Velmi často je záznam dat objemný, neorganizovaný a rozkládá se na více portálech. Názory a pocity jsou vyjádřeny různými způsoby, například typem použité slovní zásoby, kontextu psaní, využití slangů, či jiných jazykových prostředků. Díky tomu je manuální analýza zdlouhavá a téměř nemožná. S analýzou sentimentu je však využívána inovativní analýza textů a zpracování přirozeného jazyka pro extrakci a klasifikaci dat. Jakmile jsou data extrahována, jsou připravena k analýze [2][20].

2.5.2 Příprava textu

Příprava textu zahrnuje čištění extrahovaných dat před provedením analýzy. Textová příprava obvykle zahrnuje identifikaci a eliminaci netextového obsahu z textové datové sady a veškeré informace, které mohou odhalit identitu recenzentů: jméno recenzenta, místo recenzenta, datum přezkoumání. Kromě toho je také z datové sady odstraněn všechno další obsah, který není relevantní pro oblast studia [20].

2.5.3 Detekce sentimentu

Třetím krokem je detekce sentimentu. Detekce sentimentu vyžaduje hodnocení a získávání recenzí a názorů z textové sady dat pomocí výpočetních úloh. Každá věta je zkoumána z hlediska subjektivity. V datové sadě jsou uchovávány pouze věty se subjektivními výrazy. Věty, které sdělují fakta a objektivní komunikaci, jsou z další analýzy vyřazeny. Detekce sentimentu se provádí na různých úrovních. Buď jednotlivých termínů a frází, úplných větách, nebo úplném dokumentu s běžně používanými technikami, jako jsou [20]:

- **Unigramy** - jedná se o klasický přístup, kde je každý prvek reprezentován jako vektor funkce založený na frekvenci jednoho slova. To je často popisováno jako balík slov přiblížení,
- **N-Gramy** - v tomto přístupu jsou rysy dokumentu reprezentovány vícenásobnými slovy v sekvenci (např. slova v párech, trojicích), který zachytí více kontextu,
- **Lemmy** - toto zahrnuje spíše použití synonym, než doslovné výrazy. Například: lepší - dobré, nejlepší - dobré. Tato metoda údajně usnadňuje klasifikaci a usnadňuje zobecnění.
- **Negace** - toto je v podstatě rozšíření k n-gram metodám. Fráze: "líbí se mi tato kniha" a "nelíbí se mi tato kniha" by se za většiny klasifikačních technik považovaly za podobné, ale s použitou negací se tak vlastně oba termíny řadí

do opačných názorových skupin. Negaci však nelze snadno identifikovat, obzvlášť pokud se ve větě používají sarkasmy či ironie. Navíc negační výraz „NE“ vždy nezvrátí polaritu. Například ve větě: "Není divu, že je to považováno za nejlepší knihu".

- **Názorová slova** - to jsou v podstatě slova, která se používají k popisu pocitu a názorů lidí (podstatná jména, slovesa, přídavná jména, příslovce). Tato slova jsou začleněna do vektoru prvku, kde představují přítomnost slova. Tato slova jsou dobrými indikátory subjektivity v dokumentu.

Není neobvyklé najít textové věty odkazující na několik objektů, vlastností a atributů. Pomocí matematických algoritmů lze analýzu sentimentu použít k extrahování těchto objektů, vlastností a atributů a kategorizaci formulářů. To pomáhá ve fázích analýzy a zvyšuje přesnost klasifikace a sumarizace dat [20].

2.5.4 Klasifikace sentimentu

Čtvrtou etapou je klasifikace polariry, která klasifikuje každou subjektivní větu v textovém souboru dat do klasifikačních skupin. Obvykle jsou tyto skupiny zastoupeny ve dvou extrémních bodech (pozitivní, negativní; dobré, špatné: líbí se, nelíbí se). Klasifikace však může zahrnovat i více bodů, které jsou podobné bodovací škále (např. 1-5 hvězd) [20].

Klasifikace sentimentu se provádí dvěma hlavními metodami. Prostřednictvím strojového učení a slovníkových metod, které byly blíže popsány v kapitole 2.4. Metody analýzy sentimentu.

2.5.5 Prezentace výstupů

Obecným účelem analýzy je převést nestrukturovaný fragmentovaný text na smysluplné informace. Po dokončení analýzy se pro zobrazení výsledku analýzy textu použije řada konvenčních možností. Hlavním z nich je použití grafického zobrazení, jako jsou koláčové, sloupcové, či řádkové grafy. Polarita je rozdělena na barvu, frekvenci, procenta a velikost. Formát prezentace závisí na zájmech výzkumu [20].

2.6 Problémy analýzy sentimentu

Analýza sentimentu je náročná vědní disciplína, která má taky své úskalí. Odhalit sentiment textu je často náročné i pro člověka, natož pro stroj. Níže je popsáno několik problémů této analýzy.

2.6.1 Subjektivita názoru

Jak již bylo zmíněno výše, analýza sentimentu je vědní obor, který analyzuje subjektivní názory lidí vůči jednotlivým entitám. Problém nastává hned v bodě subjektivity názoru daného člověka. Například věta: „Tento byt se nachází v přízemí.“. Pro jakéhokoliv člověka, který se zrovna nesnaží koupit byt, může být tato informace naprosto bezvýznamná. Zároveň však může vyjádřit negativní sentiment, pokud stojí v inzerátu, který čte osoba, jež hledá byt s výhledem a terasou. Naopak jinak bude na věc nahlížet osoba na invalidním vozíku, která se snaží najít bydlení se snadným přístupem. Je tedy jasné, že lidé se nejsou schopni shodnout na tom, co je pozitivní, negativní, či neutrální.

Zároveň je evidentní, že je potřeba znát kontext celého textu k vytvoření si správného názoru. Například pokud žena hodnotí pračku slovy: „Je to starý model, vejde se do ní málo prádla a má větší spotřebu vody.“, bylo by možné se domnívat, že hodnotí pračku negativně. Pokud však přidá i druhou větu: „Ale na chalupu je to víc než luxus.“ dostane najednou hodnocení úplně opačný sentiment.

K problémům také dochází u frází, které mají specifický význam a současně popisují nějakou věc. Například věta: „To je bomba!“ Autor věty mohl buď vyjádřit své nadšení ustáleným spojením, či popisovat výbušninu.

2.6.2 Sarkasmus a ironie

Dalším významným problémem v souvislosti s kontextem textu je použití sarkasmu či ironie. Typická je například věta „Tak to je teda opravdu úžasné!“ Slovo úžasné je samo o sobě bráno jako velmi pozitivní výraz a v tomto případě je ještě navíc zesíleno slovem opravdu. Pokud by však byla věta myšlena ironicky, dostala by úplně opačný význam. Sarkasmus se vyskytuje spíše v politických debatách než v hodnocení produktů.

2.6.3 Falešné recenze

Jak již bylo zmíněno v úvodu, internet v mnohém změnil způsob komunikace mezi lidmi. Lidé každý den sdílejí miliony svých názorů a velice snadno tak informují a ovlivňují ostatní.

Pokud si tedy chce zákazník pořídit nové zboží, s velkou pravděpodobností si přečte hodnocení ostatních spotřebitelů, kteří již tento produkt zakoupili. Stejně tak tyto hodnocení mohou sloužit prodejcům jako zpětná vazba a možnost odhalení vad daného produktu či služby. Pozitivní hodnocení má tedy příznivý dopad na prodej výrobků, proto někteří prodejci vytváří falešná hodnocení, za cílem zvýšit pozitivní sentiment o výrobku a tím i jeho

prodej. Velmi častý je však také opačný problém, kdy výrobce šíří negativní recenze o produktech či službách konkurence.

Tyto falešné recenze se bohužel nevyskytují jen u hodnocení produktů a služeb, ale čím dál častěji se objevují v sociální a politické oblasti, což může být velice nebezpečné. Jednotlivec či skupina může prostřednictvím falešných zpráv (fake news) ovlivňovat a zmanipulovat masy.

2.6.4 Problémy slovníků

Jak již bylo zmíněno výše, k analýze sentimentu se využívají tzv. slovníky sentimentu, které obsahují seznam pozitivních a negativních slov, pomocí kterých informační program hodnotí polaritu sentimentu celkového textu. Tyto seznamy však také mají své nedostatky.

Například jednotlivá slova mohou obsahovat pozitivní i negativní sentiment. Například slovo „nepředvídatelný“ může být chápáno pozitivně, pokud je popisován film, naopak také negativně, jestliže jde o popis nového auta.

Dalším problémem je, že věta obsahující názorová slova nemusí nutně vyjadřovat sentiment. K tomuto jevu dochází často v několika typech vět. Otázky a podmíněné věty jsou dva důležité typy, např. „Můžete mi říct, která kamera Sony je dobrá?“ „Pokud v obchodě najdu dobrou kameru, koupím ji.“ Obě tyto věty obsahují názorové slovo "dobrý", ale ani jedna nevyjadřuje kladný či záporný názor na jakoukoliv konkrétní kameru.

Naopak mnoho vět bez názorových slov může také sentiment vyjadřovat. Mnohé z těchto vět jsou vlastně objektivní věty, které slouží k vyjádření některých faktických informací. Například věta: "Tato pračka používá hodně vody." implikuje negativní sentiment o pračce, protože spotřebovává mnoho zdrojů. A věta: "Po spánku na matraci po dobu dvou dnů se uprostřed vytvořila proleženina." vyjadřuje negativní názor na matraci. Obě tyto věty tak vyjadřují sentiment, ačkoliv neobsahují názorová slova.

3 ANALÝZA SENTIMENTU V NOVINOVÝCH ČLÁNCÍCH

V následující kapitole bude provedena analýza sentimentu novinových článků. Nejdříve je popsán sběr a příprava dat, poté jejich implementace do programu STATISTICA Cz, kde probíhala samotná analýza sentimentu. Následně jsou popsány výsledky analýzy sentimentu.

3.1 Sběr a příprava dat

Zdrojem pro sběr novinových článků byl server Reuters.com, což je zpravodajská a mediální divize společnosti Thomson Reuters. Reuters je největším světovým poskytovatelem multimediálních zpráv na světě. Poskytuje důvěryhodné obchodní, finanční, národní a mezinárodní novinky přímo na Reuters.com a také prostřednictvím televize Reuters [26]. Tento zdroj byl vybrán, protože umožňuje přehledné zobrazení všech novinových článků, které byly o zvolené firmě publikovány v jeden den.

Pro analýzu v této diplomové práci bylo vybráno 10 společností, u kterých proběhl sběr novinových článků v období tří let od 1. 1. 2016 do 31. 12. 2018. Za každý jeden den vznikl textový soubor, který obsahoval všechny zprávy z daného dne. Celkově tak mohlo u každé společnosti vzniknout 1096 souborů. V některý den však žádné novinové články o dané firmě nevyšly. Průměrně se počet dní obsahující text k analýze pohyboval okolo 440.

Aby bylo docíleno výběru firem s vysokým počtem novinových článků, provedla autorka sběr článků u dvaceti významných firem obchodovaných na americké burze a poté vybrala deset s největším počtem těchto dní.

Seznam dvaceti zkoumaných firem je uveden níže:

Alphabet Inc.; Amazon.com, Inc.; Apple Inc.; AT&T Inc.; Bank of America Corporation; Exxon Mobil Corporation; General Electric Company; Chevron Corporation; Intel Corporation; Johnson & Johnson; JPMorgan Chase & Co.; Microsoft Corporation; Pfizer Inc.; The Coca-Cola Company; The Home Depot, Inc; The Procter & Gamble Company; The Walt Disney Company; Verizon Communications Inc.; Walmart Inc.; Wells Fargo & Company.

V následující tabulce č. 1 je přehled dvaceti zkoumaných firem, seřazených podle četnosti výskytu dní, kdy se o dané firmě publikovalo a jejich zkratky.

Tabulka 1 - Seznam vybraných firem

Název společnosti	Zkratka	Počet publikovaných dní
Apple Inc.	AAPL	656
Amazon.com, Inc.	AMZN	614
JPMorgan Chase & Co.	JPM	448
Exxon Mobil Corporation	XOM	446
Wells Fargo & Company	WFC	399
General Electric Company	GE	374
Walmart Inc.	WMT	361
AT&T Inc.	T	355
Pfizer Inc.	PFE	354
Microsoft Corporation	MSFT	349
Alphabet Inc.	GOOG	322
Johnson & Johnson	JNJ	296
The Walt Disney Company	DIS	279
Chevron Corporation	CVX	276
Bank of America Corporation	BAC	251
Verizon Communications Inc.	VZ	247
Intel Corporation	INTC	231
The Coca-Cola Company	KO	170
The Procter & Gamble Company	PG	133
The Home Depot, Inc.	HD	71

Zdroj: vlastní zpracování

Dále bylo nutné stáhnout potřebné finanční informace o hodnotách cen akcií pro výpočet hodnot výnosu akcie. Tyto data byla stažena na portálu Yahoo Finance [10]. I tady však bylo potřeba ošetřit prázdné hodnoty a vynechat dny, kdy se na burze neobchodovalo. Bylo potřeba vybrat jen dny, kdy byla známá cena akcie k danému dni a zároveň byl o firmě publikován novinový článek.

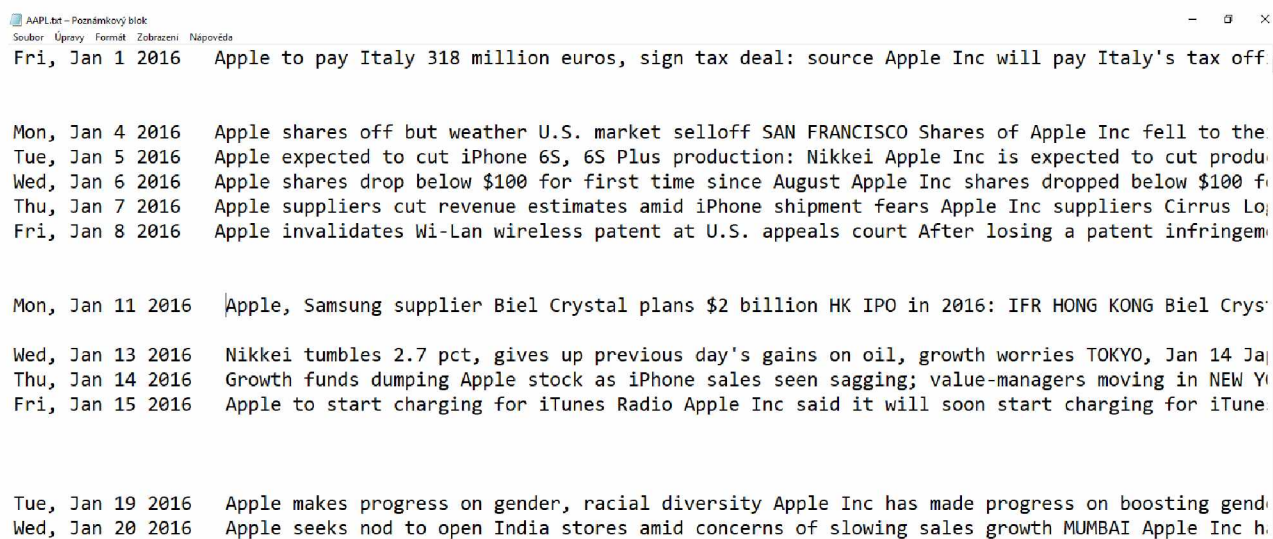
V tabulce č. 2 je zobrazen výsledný počet dní ke zpracování, očištěn od chybějících údajů. Celkem tak bylo zpracováno a připraveno k požití 3930 textových souborů.

Tabulka 2 - Výsledný počet dní

Název společnosti	Zkratka	Počet dní
Apple Inc.	AAPL	585
Amazon.com, Inc.	AMZN	552
JPMorgan Chase & Co.	JPM	424
Exxon Mobil Corporation	XOM	405
Wells Fargo & Company	WFC	376
General Electric Company	GE	348
Walmart Inc.	WMT	335
AT&T Inc.	T	327
Pfizer Inc.	PFE	339
Microsoft Corporation	MSFT	239

Zdroj: vlastní zpracování

Data byla stažena ze serveru Reuters ve formě html souborů, ze kterých byly následně extrahovány samotné články pomocí jednoduchého programu v jazyce C#. Vznikly tak textové soubory, které obsahovaly novinové články vztahující se k danému dni. Kvůli následné implementaci textových dat do tabulky programu STATISTICA Cz byly nakonec všechny textové soubory vloženy do jednoho textového souboru, kde každý řádek reprezentoval jeden den. Pokud byl tedy řádek prázdný, nebyl ten den zveřejněn žádný novinový článek o dané firmě. Ukázkou tohoto textového souboru lze vidět na obrázku č. 4.



Obrázek 4 - Ukázka textového souboru

Zdroj: vlastní zpracování

3.2 Měření sentimentu

Takto zpracovaný textový soubor byl vložen do programu STATISTICA Cz, kde byl každý řádek (čili den) jednotlivě analyzován podle zvoleného slovníku sentimentu. V této práci byly použity celkem tři slovníky sentimentu od autorů Bing Liu, Elaine Henry a kolegů Tima Loughrana a Billa McDonalda. Nástroj Text Mining programu STATISTICA Cz porovnal slova textu novinových článku se seznamy pozitivních a negativních slov zmíněných slovníků.

Tyto slovníky se od sebe liší nejen způsobem, jakým vznikly, ale i počtem slov, které obsahují. Slovník Binga Liu obsahuje 2005 pozitivních a 4775 negativních slov [15]. Je tak nejrozsáhlejším slovníkem, který byl v této práci použit, což se promítlo i v počtu nalezených slov. Slovník Tima Loughrana a Billa McDonalda (dále jen L&M) byl využit v jejich práci *When is a Liability not a Liability?* [20] a obsahuje kromě 354 pozitivních a 2355 negativních slov i seznamy neurčitých, právnických a modálně silných a slabých slov. Slovník dle Elaine Henry obsahuje nejmenší počet slov – 105 pozitivních a 85 negativních. Autorka je přesně vypsala ve své práci *Are Investors Influenced by the Way Earnings Press Releases are Written?* [12].

Rozdíly slovníků lze vidět na výsledcích analýzy v programu STATISTICA Cz. Na obrázcích 4-9 se nachází nejčastěji se vyskytující pozitivní a negativní slova v novinových článcích o firmě Apple Inc dle jednotlivých slovníků.

Na obrázku č. 5 lze vidět nejčastější pozitivní slova dle slovníku Binga Lui. Slovník vyhledal celkem 42 kořenů slov. Nejpoužívanějším slovem bylo slovo *top*, které se v novinových článcích vyskytlo celkem 149krát. Hned za ním se umístilo slovo *gain* se 146 výskyty dále následovalo slovo *boost*, které se v článcích objevilo celkem 134krát.

Seznam pozitivních slov ve slovníku L&M neobsahuje slovo *top*, proto program STATISTICA Cz vyhodnotil jako nejpočetnější pozitivní výraz slovo *gain*, následováno slovem *boost*, stejně jako v prvním slovníku, tedy i se stejnými počty výskytu. Na obrázku č. 6 si lze také všimnout, že slovník L&M vynechává výrazy jako *appeal*, *work a like*.

Documents: 1096
Words: 42 + 224

Frequency (importance/relevance measure)
 [inverse document frequency [recommended]]
 Raw Binary Logarithmic

Summary
Back
Options
Save specs

Stem / Phrase	Count	Files	Example
<input checked="" type="checkbox"/> top	149	71	
<input checked="" type="checkbox"/> gain	146	75	
<input checked="" type="checkbox"/> boost	134	69	
<input checked="" type="checkbox"/> appeal	123	56	
<input checked="" type="checkbox"/> work	96	48	
<input checked="" type="checkbox"/> like	93	58	
<input checked="" type="checkbox"/> strong	79	46	
<input checked="" type="checkbox"/> lead	60	37	

Obrázek 5 - Nejčastější pozitivní slova dle Liu

Zdroj: vlastní zpracování

Documents: 1096
Words: 15 + 61

Frequency (importance/relevance measure)
 [inverse document frequency [recommended]]
 Raw Binary Logarithmic

Summary
Back
Options
Save specs

Stem / Phrase	Count	Files	Example
<input checked="" type="checkbox"/> gain	146	75	
<input checked="" type="checkbox"/> boost	134	69	
<input checked="" type="checkbox"/> strong	79	46	
<input checked="" type="checkbox"/> win	50	25	
<input checked="" type="checkbox"/> despit	39	22	despite
<input checked="" type="checkbox"/> good	39	16	
<input checked="" type="checkbox"/> best	37	17	
<input checked="" type="checkbox"/> valuabl	37	16	valuable

Obrázek 6 - Nejčastější pozitivní slova dle Loughrana a McDonald

Zdroj: vlastní zpracování

Stem / Phrase	Count	Files	Example
<input checked="" type="checkbox"/> high	159	68	
<input checked="" type="checkbox"/> record	136	53	
<input checked="" type="checkbox"/> rise	130	76	
<input checked="" type="checkbox"/> growth	119	61	
<input checked="" type="checkbox"/> higher	111	62	
<input checked="" type="checkbox"/> strong	79	46	
<input checked="" type="checkbox"/> largest	74	51	
<input checked="" type="checkbox"/> increase	64	37	increase

Obrázek 7 - Nejčastější pozitivní slova dle Henry

Zdroj: vlastní zpracování

Dle slovníku Elaine Henry (obrázek č. 7) bylo nalezeno 16 kořenů slov, kde první čtyři nejčastější pojmy byly zcela odlišné. Byla to slova *high* (které se stalo nejčastěji vyskytovaným pozitivním slovem ve všech novinových článcích o firmě Apple Inc), *record*, *rise* a *growth*. Lze tak vidět, že ačkoliv tento slovník obsahuje několikanásobně méně slov než ostatní, je také důležitým zdrojem analýzy.

Rozdíly slovníků lze rovněž sledovat i u negativních slov. Na obrázku č. 8 jsou vidět nejčastěji vyskytovaná negativní slova při použití slovníku Binga Liu. Nejpočetnějším se stal výraz *fall*, který byl použit 142krát. Stejně tak tomu bylo i u slovníku Henry, který zároveň 142krát vyhledal i slovo *drop*, což lze vidět na obrázku č. 8. U slovníku L&M na obrázku č. 9 je znázorněn rozdíl v počtu nalezených kořenů slov, který byl u pozitivních pouze 15, kdežto u negativních 47. Je to dáno velkým rozdílem obsahu seznamů pozitivních a negativních slov.

Results: Tabulka50

Documents: 1096
Words: 48 + 372

Frequency (importance/relevance measure)
 Inverse document frequency [recommended]
 Raw Binary Logarithmic

Summary
Back
Options
Save specs

Stem / Phrase	Count	Files	Example
<input checked="" type="checkbox"/> fall	142	73	
<input checked="" type="checkbox"/> disput	99	45	dispute
<input checked="" type="checkbox"/> loss	97	53	
<input checked="" type="checkbox"/> fell	82	57	
<input checked="" type="checkbox"/> issu	82	40	issue
<input checked="" type="checkbox"/> weak	81	42	
<input checked="" type="checkbox"/> rival	77	48	
<input checked="" type="checkbox"/> slow	75	36	

Obrázek 8 - Nejčastější negativní slova dle Liu

Zdroj: vlastní zpracování

Results: Tabulka50

Documents: 1096
Words: 47 + 226

Frequency (importance/relevance measure)
 Inverse document frequency [recommended]
 Raw Binary Logarithmic

Summary
Back
Options
Save specs

Stem / Phrase	Count	Files	Example
<input checked="" type="checkbox"/> cut	128	65	
<input checked="" type="checkbox"/> disput	111	49	dispute
<input checked="" type="checkbox"/> infring	98	49	infringe
<input checked="" type="checkbox"/> loss	97	53	
<input checked="" type="checkbox"/> antitrust	96	37	
<input checked="" type="checkbox"/> warn	85	34	
<input checked="" type="checkbox"/> weak	81	42	
<input checked="" type="checkbox"/> slow	75	36	

Obrázek 9 - Nejčastější negativní slova dle LM

Zdroj: vlastní zpracování

Documents: 1096
Words: 15 + 16

Frequency (importance/relevance measure)
 Inverse document frequency [recommended]
 Raw Binary Logarithmic

Summary
Back
Options
Save specs

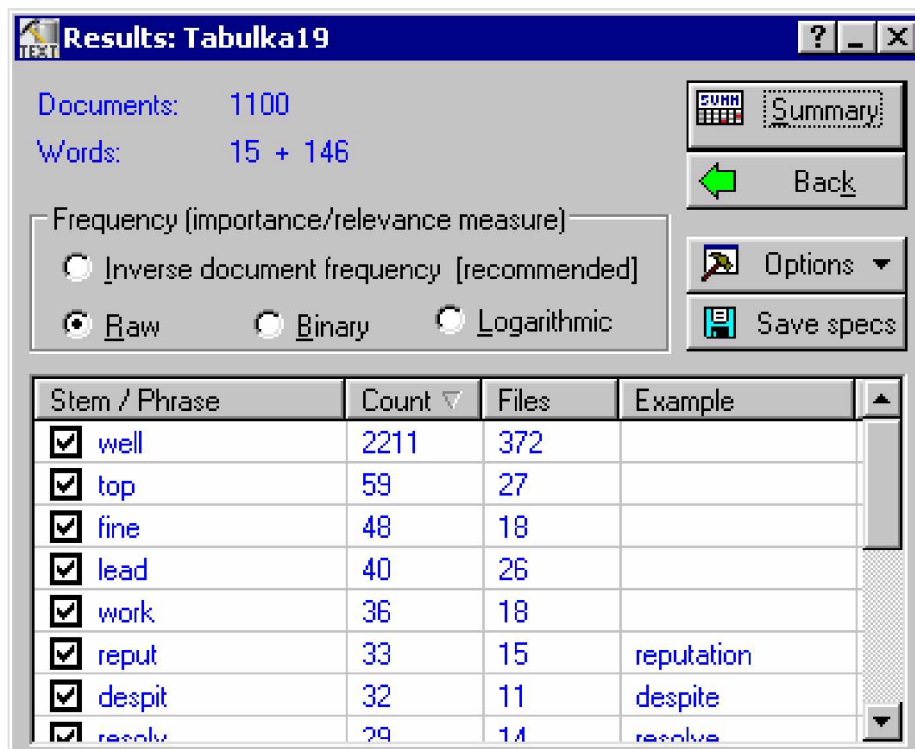
Stem / Phrase	Count	Files	Example
<input checked="" type="checkbox"/> drop	142	73	
<input checked="" type="checkbox"/> fall	142	73	
<input checked="" type="checkbox"/> lower	122	66	
<input checked="" type="checkbox"/> fell	82	57	
<input checked="" type="checkbox"/> weak	81	42	
<input checked="" type="checkbox"/> declin	62	38	decline
<input checked="" type="checkbox"/> low	54	34	
<input checked="" type="checkbox"/> challeng	45	27	challenge

Obrázek 10 - Nejčastější negativní slova dle Henry

Zdroj: vlastní zpracování

Na obrázku č. 11 lze vidět jeden z příkladů problémů slovníků sentimentu, jež byly diskutovány v kapitole 2.6.4. Problémy slovníků. Při analýze novinových článků společnosti Wells Fargo & Company bylo zjištěno nadprůměrné množství výskytu slova *well*, jelikož program vyhodnotil název společnosti jako kořen tohoto slova, což vedlo k ovlivnění výsledků výzkumu.

Výsledné počty pozitivních a negativních slov dle jednotlivých slovníků jsou znázorněny na obrázku č. 12, kde je zachycen průběh měření sentimentu společnosti Apple Inc. v programu STATISTICA Cz. V prvním sloupci je vidět načtený textový soubor novinových článků po dnech, v druhém a třetím sloupci jsou počty pozitivních a negativních slov dle slovníku Binga Liu, ve čtvrtém a pátém sloupci dle slovníku Elaine Henry a v šestém a sedmém sloupci dle slovníku L&M.



Obrázek 11 - Příklad problému slovníku sentimentu

Zdroj: vlastní zpracování

STATISTICA Cz - [Data: Tabulka19]

Menu: Soubor, Domů, Upravit, Zobrazit, Formát, Statistika, Data mining, Grafy, Nástroje, Data

Data mining options: Neuronové sítě, Machine Learning, GAM, Učení, IC Analysis, Optimal Binning, Shlukování/Skupiny, Text Mining, Association, Link Analysis, Pravidlo extra

	1 Novinové články	2 Pozit-Liu	3 Negat-Liu	4 Pozit-Henry	5 Negat-Henry	6 Pozit-Lougran+McDonald	7 Negat-Lougran+McDonald	8 Prom
1	Fri, Jan 1 2016 Apple to pay Italy	3	6			1		4
2								
3								
4	Mon, Jan 4 2016 Apple shares off	5	2	1	3	1		4
5	Tue, Jan 5 2016 Apple expected to	23	12	1	1	1		18
6	Wed, Jan 6 2016 Apple shares dro	31	10	5	6	2		18
7	Thu, Jan 7 2016 Apple suppliers c	10	17	2	3			23
8	Fri, Jan 8 2016 Apple invalidates V	21	16	3	5	2		13
9								
10								
11	Mon, Jan 11 2016 Apple, Samsung	3				1		1
12								
13	Wed, Jan 13 2016 Nikkei tumbles	3	5	1	2	2		4
14	Thu, Jan 14 2016 Growth funds dui	6	9	6		2		4
15	Fri, Jan 15 2016 Apple to start cha	3	1					2
16								
17								
18								
19	Tue, Jan 19 2016 Apple makes pro	7		4		6		4
20	Wed, Jan 20 2016 Apple seeks no	8	10	9		2		10
21								
22	Fri, Jan 22 2016 Fewer orders at A	13	8	1	3	1		12
23	Sat, Jan 23 2016 Apple's Zadeski	1	1					

Obrázek 12 - Počty pozitivních a negativních slov dle jednotlivých slovníků

Zdroj: vlastní zpracování

3.3 Výsledky měření sentimentu

Jak již bylo zmíněno výše, obrázek č. 12 zobrazuje průběh měření sentimentu společnosti Apple Inc. a výskyt pozitivních a negativních slov. Tyto výsledky byly použity k výpočtu poměrů sentimentu za jednotlivé dny a to pro každou společnost. K původním sloupcům tak přibyly další hodnoty, které jsou zobrazeny na obrázku č. 13. K těmto výpočtům došlo v programu Excel.

Výsledný poměr sentimentu byl vypočítán pomocí vzorce:

$$\text{poměr sentimentu} = \frac{P-N}{P+N} \quad (2)$$

kde P je počet pozitivních slov a N počet negativních slov k danému dni. Takto vyjádřený poměr mohl nabývat hodnot $\langle -1; 1 \rangle$, kde -1 vyjadřuje 100% negativní sentiment a hodnota 1 zobrazuje 100 % pozitivní sentiment.

	A	B	C	D	E	F	G	H	I	J	K	L	M
	Datum	POZ-Liu	NEG-Liu	Poměr-Liu	POZ-McDonald	NEG-McDonald	Poměr-McDonald	POZ-Henry	NEG-Henry	Poměr-Henry			
2	1.1.2016	3	6	-0,33	0	4	-1,00	0	1	-1,00			
3	4.1.2016	5	2	0,43	1	4	-0,60	1	3	-0,50			
4	5.1.2016	23	12	0,31	1	18	-0,89	1	1	0,00			
5	6.1.2016	31	10	0,51	2	18	-0,80	5	6	-0,09			
6	7.1.2016	10	17	-0,26	0	23	-1,00	2	3	-0,20			
7	8.1.2016	21	16	0,14	2	13	-0,73	3	5	-0,25			
8	11.1.2016	3	0	1,00	1	1	0,00						
9	14.1.2016	3	5	-0,25	2	4	-0,33	1	2	-0,33			
10	15.1.2016	6	9	-0,20	2	4	-0,33	6	0	1,00			
11	16.1.2016	3	1	0,50	0	2	-1,00						
12	19.1.2016	7	0	1,00	6	4	0,20	4	0	1,00			
13	20.1.2016	8	10	-0,11	2	10	-0,67	9	0	1,00			
14	22.1.2016	13	8	0,24	1	12	-0,85	1	3	-0,50			
15	23.1.2016	1	1	0,00									
16	25.1.2016	4	1	0,60	2	1	0,33	3	1	0,50			
17	26.1.2016	22	14	0,22	3	15	-0,67	4	1	0,60			
18	27.1.2016	12	20	-0,25	2	14	-0,75	11	20	-0,29			
19	28.1.2016	26	23	0,06	5	24	-0,66	2	8	-0,60			
20	29.1.2016	22	15	0,19	4	12	-0,50	2	4	-0,33			
21	1.2.2016	19	3	0,73	10	0	1,00	8	0	1,00			
22	2.2.2016	34	8	0,62	18	12	0,20	5	2	0,43			

Obrázek 13 - Poměry naměřeného sentimentu

Zdroj: vlastní zpracování

Tabulka 3 - Průměr naměřeného sentimentu za sledované období

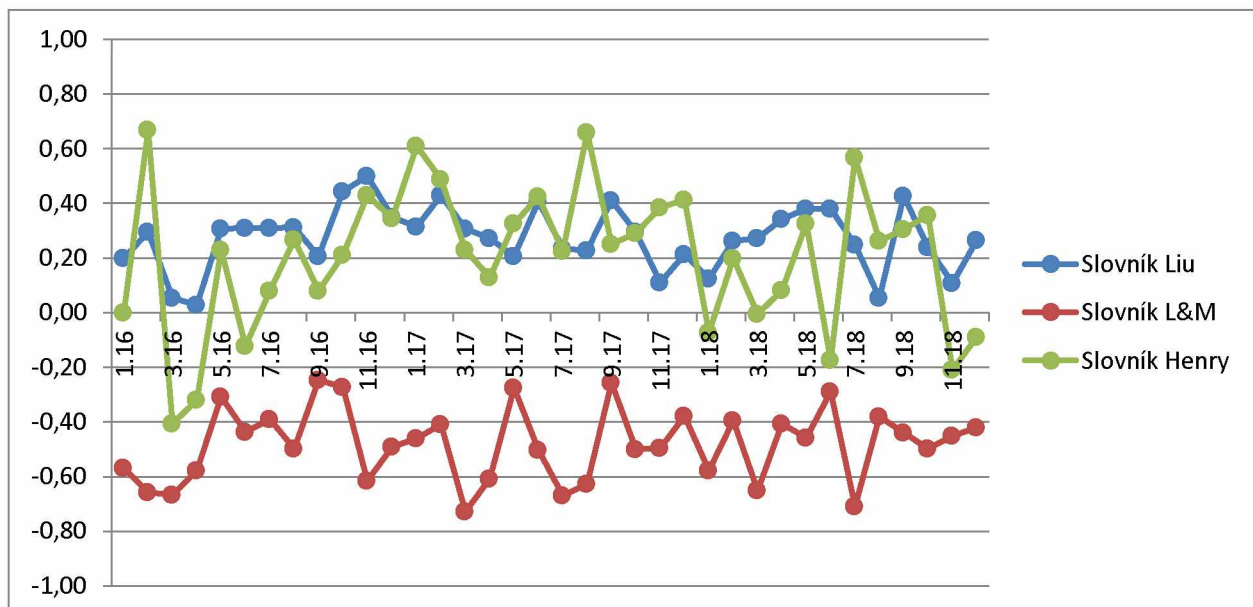
Zkratka	Poměr - Liu	Poměr - L&M	Poměr - Henry
AAPL	0,27	-0,48	0,19
AMZN	0,37	-0,33	0,4
JPM	0,35	-0,33	0,35
XOM	0,33	-0,56	0,33
WFC	0,39	-0,54	0,27
GE	0,46	-0,34	0,2
WMT	0,36	-0,3	0,53
T	0,45	-0,51	0,3
PFE	0,33	-0,21	0,24
MSFT	0,34	-0,31	0,45

Zdroj: vlastní zpracování

Jelikož byly hodnoty sentimentu měřeny k jednotlivým dnům a vzniklo tak vysoké množství dat, jsou výsledky prezentovány v průměru za určité období. V tabulce č. 3 lze vidět poměry sentimentu dle jednotlivých slovníků pro všech 10 firem. Ty jsou zároveň zprůměrovány za celé období tří let. Lze na nich sledovat, že všechny hodnoty u slovníku L&M jsou záporné, což značí nalezený negativní sentiment ve zprávách dle tohoto slovníku. Vysvětlení spočívá především ve velkém rozdílu počtu výrazů v seznamech pozitivních a negativních slov u tohoto slovníku, jak již bylo zmíněno výše.

Na následných grafech (obrázky 14 až 23) jsou vyobrazeny hodnoty sentimentu pro každou firmu zvlášť. Na ose Y jsou naneseny hodnoty poměru sentimentu, které dosahují hodnot $<-1;1>$, na ose X potom název měsíce, za který byly hodnoty zprůměrovány. Stejně jako v předchozí tabulce lze u všech grafů sledovat větší počet nalezených negativních slov u slovníku L&M.

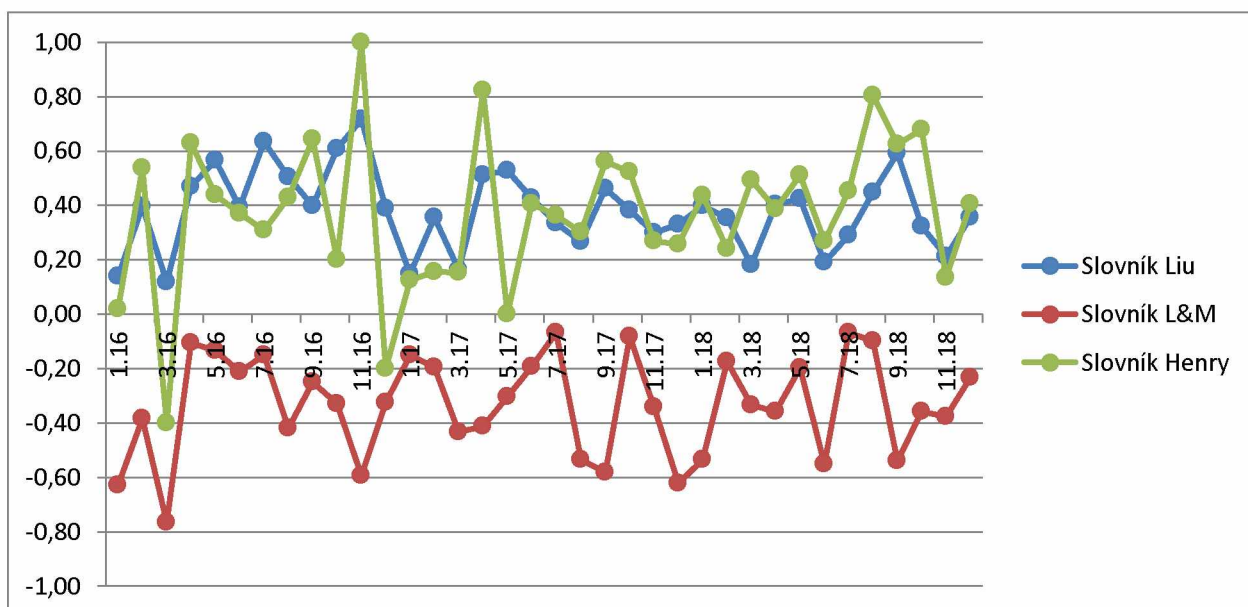
Na prvním grafu (obrázek č. 14) jsou zaznačeny průměrné měsíční hodnoty sentimentu pro firmu Apple Inc. Tato společnost měla největší počet nalezených novinových článků. Ty se vyskytovaly více než každý druhý den, přičemž jich bylo za den i několik. Z tohoto důvodu měla tato firma nejvyšší počet nalezených pozitivních a negativních slov. Dá se tedy předpokládat, že měření u této společnosti bude mít největší vypovídající hodnotu. Lze pozorovat, že hodnoty naměřené slovníky Liu a Henry se pohybují v podobných hodnotách, ačkoliv podobnost vývoje sentimentu není tak patrná jako například u následující společnosti.



Obrázek 14 - Průměrné měsíční hodnoty sentimentu firmy AAPL

Zdroj: vlastní zpracování

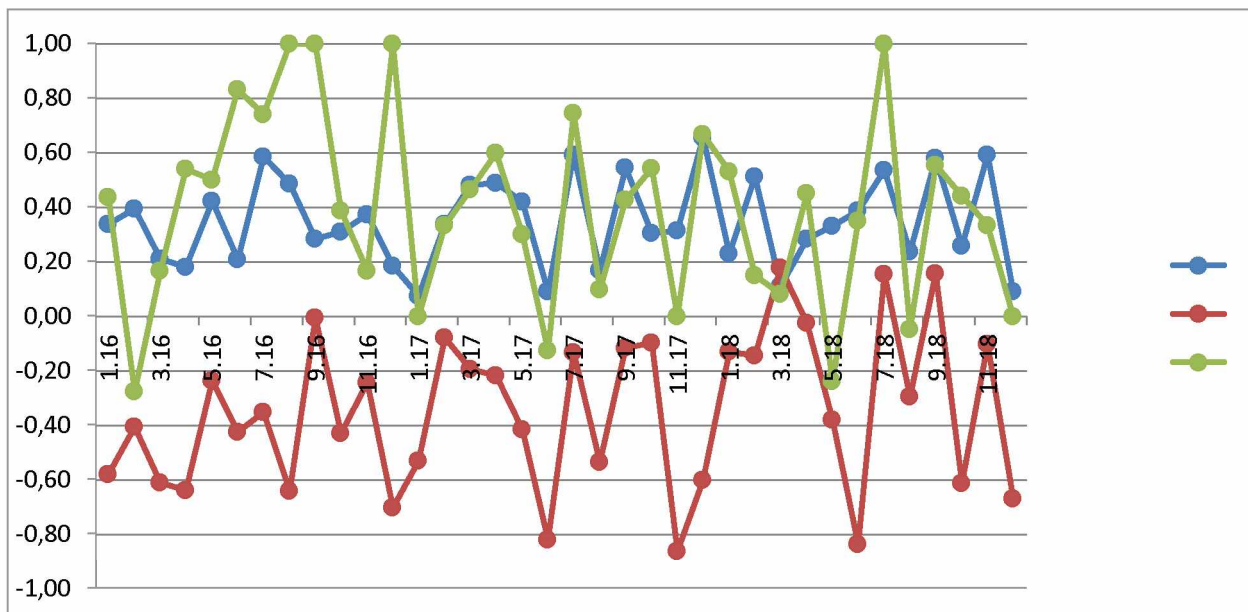
Na obrázku č. 15 lze spatřit grafické znázornění výsledků sentimentu u firmy AMAZON.com, Inc. Zhruba od půlky sledovaného období, tedy června 2017, je vidět velice podobný vývoj sentimentu u slovníku Liu a Henry. V některých částech se dokonce téměř překrývají. Tato společnost měla také velký počet zdrojových dat.



Obrázek 15 - Průměrné měsíční hodnoty sentimentu firmy AMZN

Zdroj: vlastní zpracování

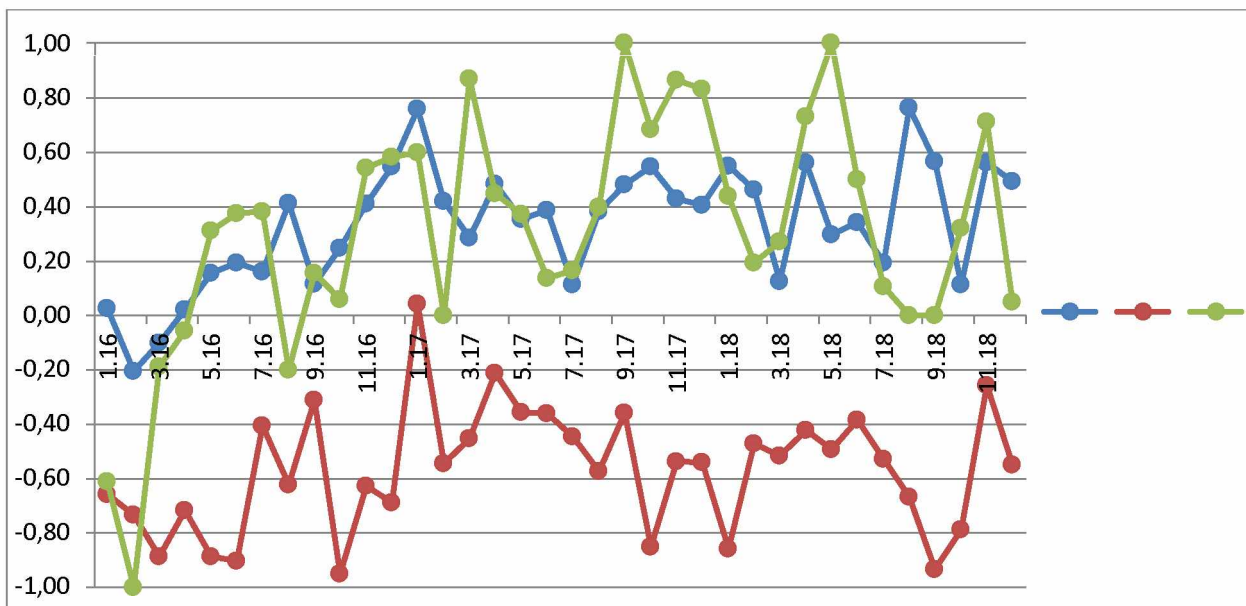
Na dalším grafu (obrázek č. 16) je zobrazen vývoj sentimentu u společnosti JPMorgan Chase & Co. Stejně jako v předchozím případě si lze všimnout téměř stejného vývoje u slovníku Liu a Henry od začátku roku 2017 až do jeho konce. A ačkoliv se slovník LM nachází v negativních hodnotách, lze sledovat korespondující vývoj sentimentu jako u ostatních slovníků.



Obrázek 16 - Průměrné měsíční hodnoty sentimentu firmy JPM

Zdroj: vlastní zpracování

V grafu průměrných měsíčních hodnot sentimentu společnosti Exxon Mobil Corporation (zobrazených na obrázku č. 17) si lze všimnout dvou zajímavých částí grafu. V srpnu 2016 a v srpnu 2018 lze vidět výrazný pokles sentimentu slovníků LM a Henry oproti růstu ve slovníku Liu. Ve většině případů spolu korespondují spíše slovníky Liu a Henry, kdežto v tomto případě můžeme sledovat celkovou podobnost právě mezi slovníky Liu a LM.

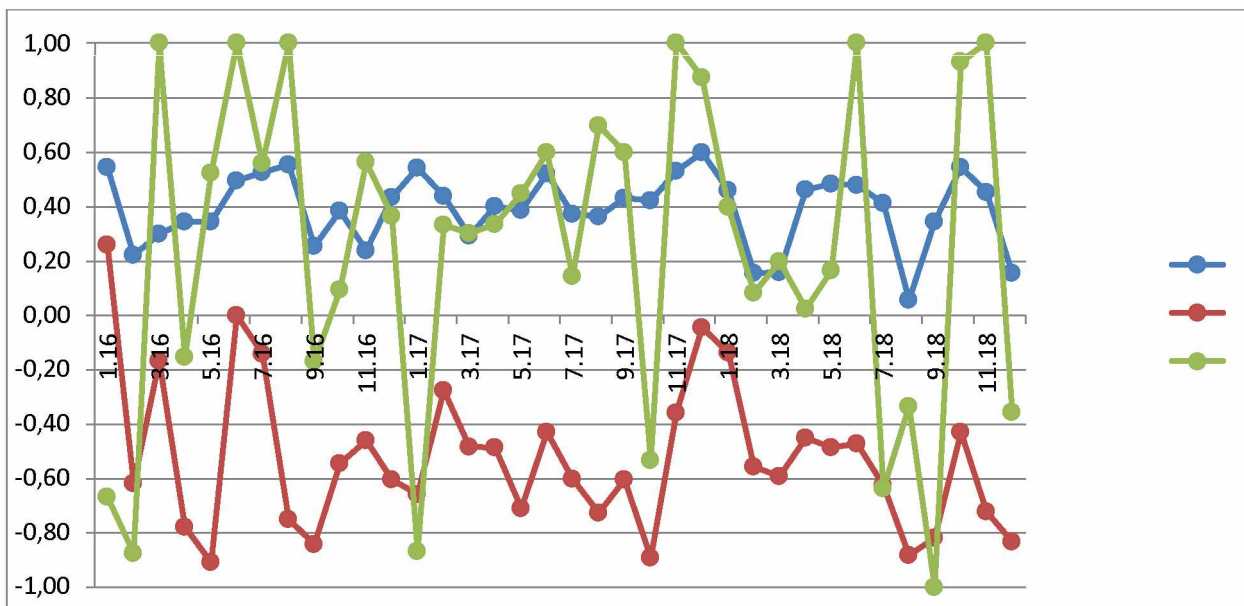


Obrázek 17 - Průměrné měsíční hodnoty sentimentu firmy XOM

Zdroj: vlastní zpracování

Na obrázku č. 18 je graficky znázorněn jeden z problémů analýzy sentimentu. Slovníky LM i Henry, obsahují v seznamu pozitivních slov slovo *well*, které tak vyhledávají ve všech novinových článcích, v tomto případě u společnosti Wells Fargo & Company. Hodnotící program tak přiřadil ke každému článku takřka vždy minimálně jedno pozitivní slovo, neboť téměř v každém článku se nachází název společnosti. Nejvíce si toho lze všimnout u slovníku Henry, protože tento slovník obsahuje nízký počet pozitivních i negativních slov (seznam negativních slov obsahuje pouze 85 pojmů oproti 2355 ve slovníku L&M). Často se tak především u krátkých zpráv stalo, že pro jednotlivý den byly nalezeny jen pozitivní slova a žádné negativní, což způsobilo výsledný poměr 1. Je možné vidět, že tento případ nastal ve více případech, jelikož hodnoty v grafu jsou uváděny průměrně za měsíc, což znamená, že všechny zprávy v daný měsíc mají poměr 1. Zajímavá je také hodnota v září 2018, kdy sentiment nabývá naopak hodnoty -1. Po bližším prozkoumání textových souborů bylo zjištěno, že daný měsíc sice obsahoval 7 dní, kdy byly publikovány články o této firmě, avšak slovník Henry detekoval názorová slova pouze v jednom dni. V tento den vyšel článek, který se zabýval soudním sporem, který se týkal také okrajově této společnosti, což však vplynulo až po přečtení z kontextu konkrétního článku. Program tedy vyhodnotil pouze negativní slova, což způsobilo již zmíněnou odchylku průměrných hodnot.

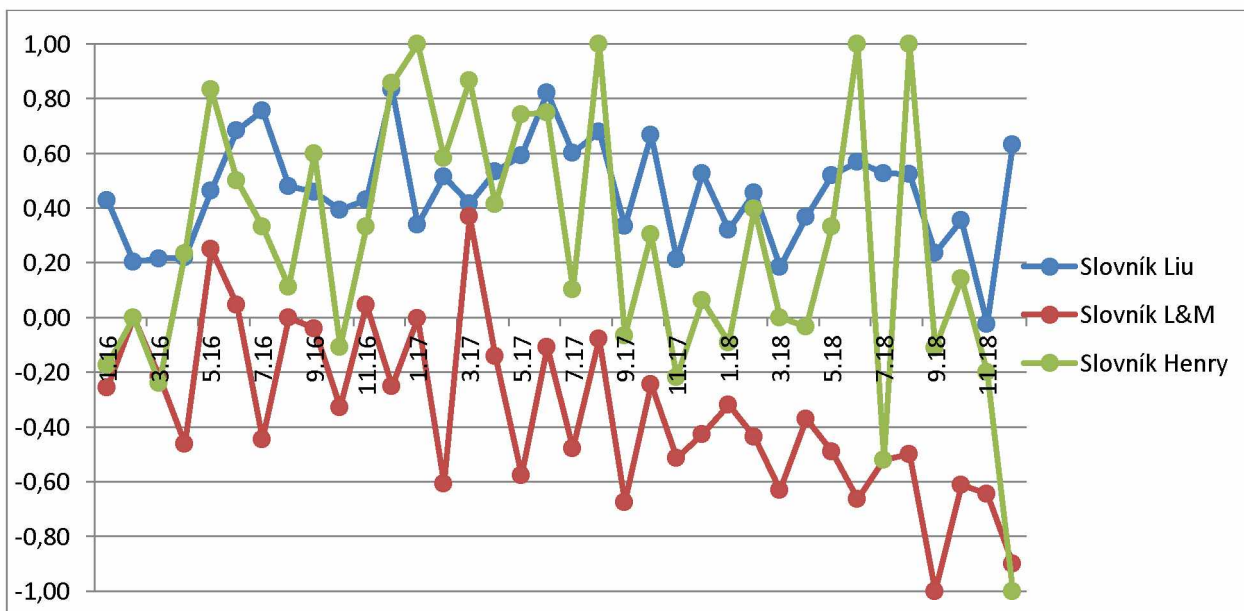
Je také možné sledovat zvýšenou podobnost vývoje sentimentu u slovníků Henry a LM od začátku do konce roku 2018.



Obrázek 18 - Průměrné měsíční hodnoty sentimentu firmy WFC

Zdroj: vlastní zpracování

Na dalších grafech (obrázky 19 až 23) se začíná projevovat vliv klesajícího počtu článků o jednotlivých firmách. Společnost General Electric Company (jejíž vývoj je znázorněn na obrázku č. 19) sice stále dosahuje počtu okolo deseti dní za měsíc, nicméně klesají počty článků za jednotlivé dny, čímž se zmenšuje počet měřených dat, který může ovlivnit poměry sentimentu a častěji tak dosahovat extrémních hodnot. Lze tak sledovat větší výkyvy sentimentu v jednotlivých měsících.

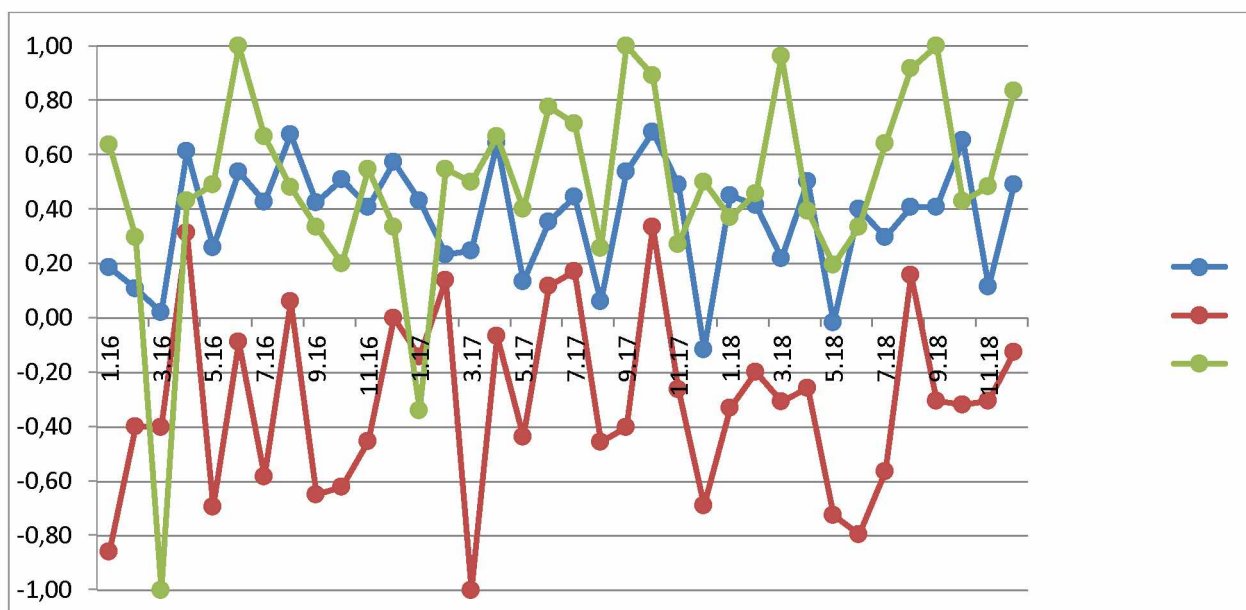


Obrázek 19 - Průměrné měsíční hodnoty sentimentu firmy GE

Zdroj: vlastní zpracování

V grafu společnosti Walmart Inc. si lze všimnout významného nárůstu sentimentu u všech tří slovníků v září 2017 a naopak poklesu na jaře 2018. Po podrobnějším prozkoumání textových souborů byly v září nalezeny novinové články zmiňující investice do internetového prodeje, například spuštění nakupování hlasem. Tyto články na autorku působily skutečně pozitivním dojmem. V květnu 2018 bylo mimo jiné zveřejněno několik článků o ztrátách zisku společnosti, způsobených právě v důsledku těchto investic.

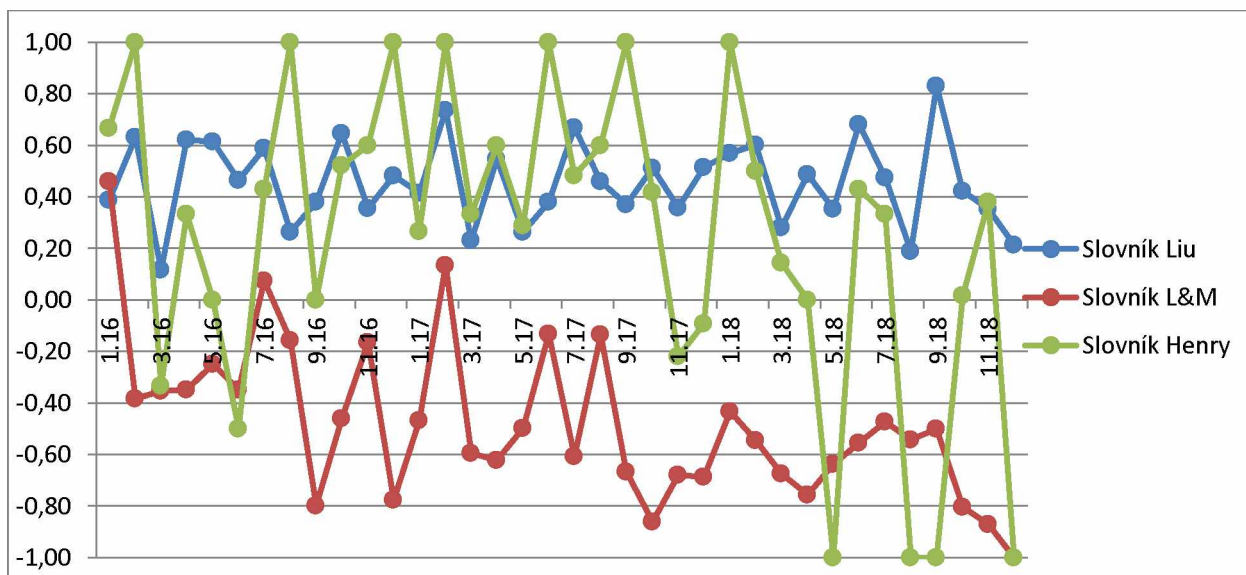
Mimo to je také zajímavé zmínit, že v tomto případě dosahují hodnoty sentimentu naměřené slovníkem LM pozitivních hodnot až v sedmi případech.



Obrázek 20 - Průměrné měsíční hodnoty sentimentu firmy WMT

Zdroj: vlastní zpracování

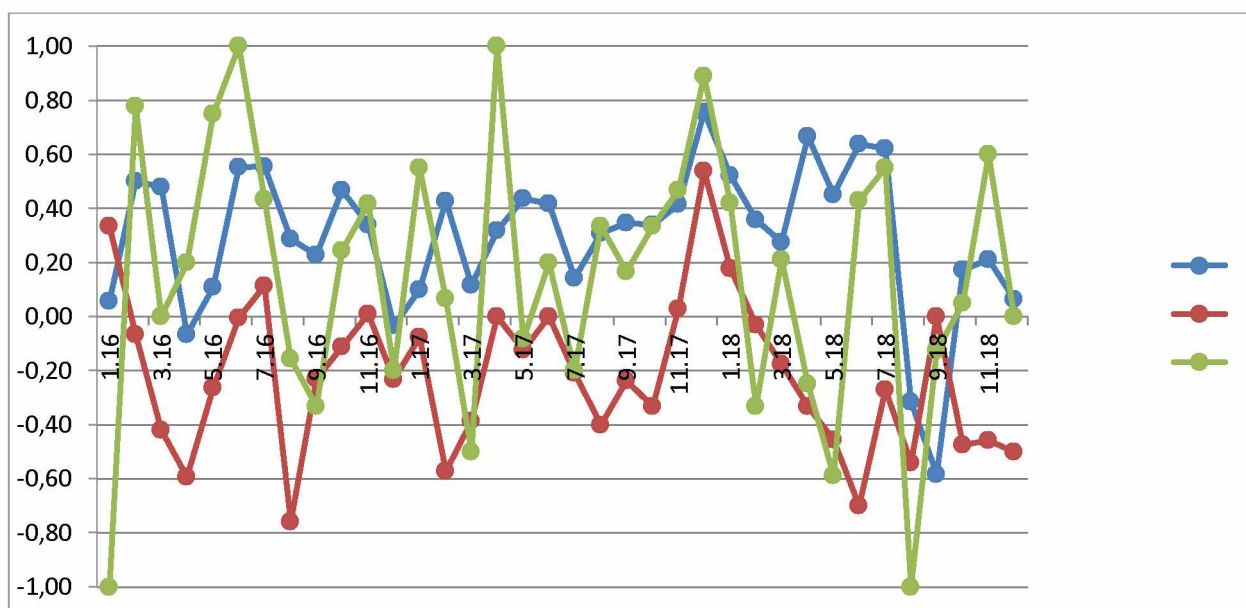
U firmy AT&T Inc, jež je graficky znázorněna na obrázku č. 21 lze vidět již výše zmíněný efekt úbytku počtu novinových článků během jednoho dne, což se projevuje nízkým počtem nalezených sentimentálních slov převážně u slovníku Henry, kdy skoro polovina všech měsíčních průměrů dosahuje krajních hodnot 1 nebo 0.



Obrázek 21 - Průměrné měsíční hodnoty sentimentu firmy T

Zdroj: vlastní zpracování

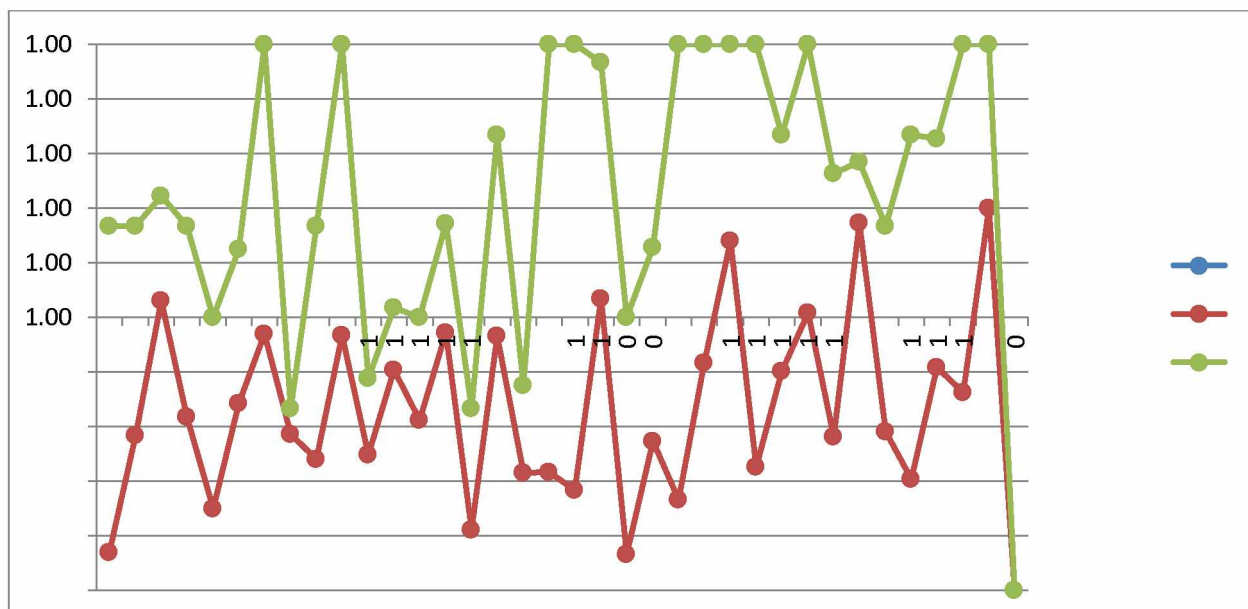
Na obrázku č. 22 jsou graficky znázorněny průměrné měsíční hodnoty sentimentu společnosti Pfizer Inc. V prosinci roku 2017 si lze všimnout významného růstu všech tří naměřených poměrů sentimentu. Slovník LM dokonce dosáhl v tomto bodě svého maxima hodnotou 0,54.



Obrázek 22 - Průměrné měsíční hodnoty sentimentu firmy PFE

Zdroj: vlastní zpracování

O společnosti Microsoft Corporation bylo publikováno nejméně dní ze zkoumaného období. Počet zdrojových dat byl až dvakrát menší než u společnosti Apple Inc. Na grafu (obrázek č. 23) tak lze sledovat větší výkyvy v hodnotách sentimentu a časté krajní hodnoty u slovníku Henry, což už bylo diskutováno výše.



Obrázek 23 - Průměrné měsíční hodnoty sentimentu firmy MSFT

Zdroj: vlastní zpracování

V následující kapitole budou tyto naměřené hodnoty využity k posouzení vlivu sentimentu na abnormální výnos akcie.

4 VLIV EMOCÍ NA ABNORMÁLNÍ VÝNOS AKCIE

Cílem této diplomové práce je potvrdit či vyvrátit vliv emocí z novinových článků na abnormální výnos akcie. Tento vliv bude zjišťován pomocí nejčastěji využívané statistické metody - regresní analýzy.

Zdrojem hodnot sentimentu pro tuto analýzu budou výsledky měření minulé kapitoly. Hodnoty abnormální akcie je potřeba nejprve vypočítat.

4.1 Výpočet abnormálního výnosu akcie

Jak bylo již uvedeno v první kapitole, abnormální výnos akcie se vypočítá dle vzorce:

$$AR_{it} = R_{it} - E(R_{it}) \quad (3)$$

kde AR_{it} je abnormální výnos firmy i za den t , R_{it} je výnos akcie firmy i za den t , a $E(R_{it})$ je očekávaný výnos trhu firmy i za den t . Abnormální výnosnost investice tak může být kladná nebo záporná.

Finanční data jednotlivých akciových titulů byla čerpána na portálu Yahoo Finance [15][10], odkud byly staženy ceny akcií pro každý den zkoumaného období během let 2016 až 2018. Z těchto cen byl vypočten výnos akcií k následujícímu dni pomocí vzorce:

$$R_{t+1} = \frac{P_{t+1} - P_t}{P_t} \quad (4)$$

Kde R_{t+1} je míra akciového výnosu následujícího dne, P_{t+1} je cena akcie následujícího dne a P_t je cena současného dne. Jedná se tak o velikost změny tržní ceny cenného papíru za dobu jeho držby, dělenou jeho původní nákupní cenou.

Ukazatelem vývoje daného trhu jako celku je burzovní index. Slouží ke sledování vývoje trhu v čase a jeho vývojovým tendencím. Každý burzovní či mimoburzovní trh má svůj vlastní index za jednotlivé druhy CP obchodovaných na trhu. Pro tuto diplomovou práci byl vybrán index Nasdaq Composite [15].

Index Nasdaq, neboli Nasdaq Composite, sleduje vývoj více než 3300 akcií, které se obchodují na americké burze NASDAQ. Spolu s indexem Dow Jones a S&P 500 je třetím nejvíce sledovaným indexem v USA. Vzhledem k tomu, že se na burze NASDAQ obchodují z velké části technologické tituly, je i vývoj ceny indexu Nasdaq Composite silně ovlivněn právě vývojem technologického sektoru [15].

Nasdaq Composite je vážený index a jeho váha se odvíjí od tržní kapitalizace společnosti. Největší vliv na vývoj ceny indexu mají společnosti, jejichž tržní kapitalizace (součin počtu a hodnoty akcií) je nejvyšší. Index je kalkulován nepřetržitě v průběhu obchodního dne, ovšem je aktualizován jednou za sekundu. Konečná hodnota je oznámena každý obchodní den v 15:16 hod [15].

Ze získaných hodnot tohoto indexu byla vypočítána míra výnosu trhu stejně jako u jednotlivých akciových titulů. Následně od sebe byly tyto hodnoty odečteny, čímž došlo ke zjištění možného vzniku abnormálního akciového výnosu. Všechny tyto výpočty probíhaly v programu Excel. Pokud abnormální akciový výnos vznikl, byl zaznačen číslem 1, pokud ne byla zaznamenána hodnota 0.

Všechny hodnoty byly vypočítány pro AR_{t+1} , AR_{t+3} a AR_{t+5} , což znamená s jednodenním, třídenním a pětidenním zpožděním.

4.2 Regresní analýza

Metody regresní analýzy jsou využívány v situacích, kdy nás zajímá závislost určité kvantitativní (spojité) proměnné na jedné nebo více dalších kvantitativních (spojitých) proměnných, tzv. regresorech. Předem je dáno, která proměnná je nezávislá (vysvětlující) a která je závislá (vysvětlovaná). Cílem regresní analýzy je popsat tuto závislost pomocí vhodného matematického modelu [27].

Dle počtu nezávisle proměnných rozlišujeme modely jednoduché regrese a vícenásobné regrese. Jednoduchá regrese popisuje závislost vysvětlované proměnné na jednom regresoru. Naproti tomu vícenásobná regrese řeší situaci, kde závisle proměnná závisí na více než jednom regresoru. Podle typu regresní funkce pak lze dále rozlišit modely lineární a nelineární[27]. V této diplomové práci byl použit model logistické regrese.

Logistická regrese je označení metody matematické statistiky zabývající se problematikou odhadu pravděpodobnosti nějakého jevu (závisle proměnné) na základě určitých známých skutečností (nezávisle proměnných), které mohou ovlivnit výskyt jevu. Událost, zda zkoumaný jev nastal, se modeluje pomocí náhodné veličiny, která nabývá hodnoty 0, pokud jev nenastal, nebo 1, pokud jev nastal. O náhodné veličině, která nabývá dvou hodnot 0 a 1 se říká, že má alternativní rozdělení. Metoda logistické regrese předpokládá, že za podmínek, které určuje vektor x bude náhodná veličina $Y(x)$ rovna 1 s pravděpodobností, jejíž závislost na x můžeme vyjádřit pomocí tzv. logistické funkce, což zapisujeme jako [20]:

$$P[Y(x) = 1] = \frac{\exp(\beta'x)}{1+\exp(\beta'x)} \quad (5)$$

kde vektor β je vektorem neznámých parametrů.

Odhadem vektoru β se tedy odhaduje i hledaná pravděpodobnost výskytu zkoumaného jevu (za předpokladu parametrizace logistickou funkcí). Vektor x se obvykle bere ve tvaru $(1, z)'$. Složka β_0 pak určuje vliv tzv. absolutního členu. Skutečnost, že pravděpodobnost výskytu jevu nezávisí na námi zkoumaných nezávislých proměnných (tj. $\beta_1 = 0, \dots, \beta_k = 0$) znamená, že se dá vyjádřit nezávisle na z ve tvaru [20]:

$$P[Y(x) = 1] = P[Y(1z)' = 1] = \frac{\exp(\beta'x)}{1+\exp(\beta'x)} \quad (6)$$

K určení, zda by bylo možné predikovat abnormální akciové výnosy na základě analýzy sentimentu byla v této diplomové práci využita logistická regresní analýza právě z důvodu binárního rozdělení závislé proměnné.

Regresní model, kde roli nezávislých proměnných hrál počet slov v novinových článcích během jednoho dne a informace poměrů sentimentu z analýzy emocí, se snažil predikovat, jak v daném období bude reagovat abnormální výnos akcie.

Všechny výpočty probíhaly v programu STATISTICA Cz.

Při posouzení 3 slovníků sentimentu, 3 typů AR a 10 firem bylo získáno 90 regresních modelů. Na obrázku č. 24 lze vidět výsledný model pro firmu Apple, Inc. při zvolení slovníku Liu a AR_{t+1} . Pro lepší přehlednost byla vytvořena tabulka se všemi výsledky pro každou firmu. Tyto tabulky jsou obsahem přílohy diplomové práce.

		AR t+1 - Test všech efektů (AAPL) Rozdělení : BINOMICKÉ, Linkující funkce: LOGIT Modelovaná pravděpodobnost, že AR t+1 = 0		
Efekt	Stupně volnosti	Wald. Stat.	p	
Abs. člen	1	5,032122	0,024881	
Poměr-Liu	1	1,181386	0,277074	
Počet slov ve zprávě	1	1,759844	0,184644	

Obrázek 24 - Ukázka modelu logistické regrese pro firmu AAPL

Zdroj: vlastní zpracování

Ze všech 90 modelů bylo nalezeno jen 7 případů, které mělo p-hodnotu poměru emočních dat menší než 0,05. Tyto případy se neděly u konkrétních AR (např. AR_{t+3}), ale byly nezávisle rozprostřeny. Vliv délky zprávy na regresor se prokázal ještě nižší (5 případů p-hodnoty $< 0,05$). Tyto případy jsou v tabulkách přiložených v příloze vyznačeny červeně.

Výsledky tak nepotvrzují hypotézu, že existuje jasný vliv a je možné na základě výše zmíněných nezávislých proměnných abnormální akciové výnosy predikovat. Lze proto vyvodit, že v určitých případech je možné, že tyto data ovlivňují výsledek, ale nelze je samotné použít k spolehlivé predikci výsledků vzniku abnormálních akcií, protože problematika je výrazně složitější a bylo by nutné do regresního modelu zahrnout další proměnné, které též dílčím způsobem ovlivňují chování na kapitálových trzích.

ZÁVĚR

Analýza sentimentu je metoda, která zkoumá subjektivní lidské názory. Je to část analýzy textu, která se využívá v široké škále odvětví. V dnešní době neustále se rozvíjejícího vlivu internetu vzniká neskutečné množství snadno dostupného textu, ve kterém není lehké se zorientovat. Analyzováním tohoto textu lze získat cenné informace v závislosti na odvětví, které je zkoumáno. Porozumění lidským postojům, názorům a emocím v psaném textu tak může pomoci nejen obchodníkům, kteří mohou na základě znalostí zákaznickových preferencí provést změny u svých produktů a služeb, ale například i politickým stranám, které mohou přizpůsobit svůj politický program dle přání voličů. V neposlední řadě se analýza sentimentu využívá také ve finanční sféře. Obsahem této diplomové práce je právě analýza vlivu sentimentu na vývoj podniků na kapitálových trzích. Tento vývoj je hodnocen na základě abnormálního akciového výnosu.

V první části práce bylo provedeno teoretické shrnutí problematiky. V první kapitole byl popsán úvod do akciových trhů. Byly vysvětleny pojmy jako akcie, burza a burzovní index. Následoval popis samotného abnormálního akciového výnosu a současné přístupy k jeho predikci.

Cílem druhé kapitoly bylo přiblížit problematiku analýzy sentimentu. Nejdříve je obecně popsán její význam a definice, následně je objasněn pojem názor, jeho druhy a prvky. Další částí této kapitoly jsou dva hlavní přístupy analýzy emocí v textových dokumentech tedy slovníková metoda a metoda strojového učení. V praktické části této práce bylo využito všeobecné metodologie analýzy sentimentu, na jejíž popis navazuje souhrn problémů a nedostatků analýzy sentimentu.

Třetí kapitola obsahuje sběr a zpracování dat, u kterých byl hodnocen sentiment. Novinové články ze zkoumaného období tří let (1. 1. 2016 – 31. 12. 2018) byly staženy ze serveru Reuters.com. Celkem bylo zkoumáno deset společností, jejichž akcie jsou obchodovány na americké burze. Jejich výběr je vysvětlen v úvodu této kapitoly. Takto stažená data ve formě textových souborů bylo potřeba očistit a připravit pro vstup do programu STATISTICA Cz, ve kterém byly zjišťovány hodnoty sentimentu pomocí tří slovníků sentimentu obsahujících seznam pozitivních a negativních slov. Celkem byly použity slovníky od třech autorů - Bing Liu, Elaine Henry a dvojice Loughran a McDonald.

V poslední části práce došlo k objasnění, zda jsou informace o sentimentu dostatečně spolehlivé pro predikci akciového vývoje. K tomuto zjištění byla využita regresní analýza, konkrétně logistická regrese.

Výsledky nepotvrdily hypotézu, že existuje jasný vliv mezi naměřenými hodnotami sentimentu a abnormálním výnosem akcií. Lze proto vyvodit, že v určitých případech je možné, že tyto data ovlivňují výsledek, ale nelze je samotná použít k spolehlivé predikci výsledků vzniku abnormálních výnosů, protože daná problematika je výrazně složitější. Výnosy akcií totiž ovlivňují jiné důležitější faktory, které by bylo potřeba do analýzy zahrnout.

POUŽITÁ LITERATURA

- [1] Abnormal Return Definition & Example | InvestingAnswers. Financial Dictionary, Calculators & Educational Articles [online] [cit. 2019-06-19]. Dostupné z: <https://investinganswers.com/dictionary/a/abnormal-return>
- [2] Akciový Trh: Význam a definice. Capital.com [online]. [cit. 2019-06-17]. Dostupné z: <https://capital.com/cs/akciovvy-trh-definice>
- [3] Burzovní index. Akcie.cz [online] [cit. 2019-06-17] Dostupné z: <https://www.akcie.cz/slovník/burzovni-index>
- [4] Burzovní indexy: 10 nejdůležitějších akciových indexů. Broker LYNX [online]. [cit. 2019-06-17] Dostupné z: <https://www.lynxbroker.cz/vzdelavani/akciove-indexy/>
- [5] Burzy a mimoburzovní trhy. Peníze.cz - Půjčky, Kurzy měn, Akcie, Hypotéky, Bydlení, Daně. [online]. [cit. 2019-06-17]. Dostupné z: <https://www.penize.cz/15864-burzy-a-mimoburzovni-trhy>
- [6] Co je to akcie. Investice.cz. [online]. [cit. 2019-05-14]. Dostupné z: <https://www.investice.cz/co-je-akcie/>
- [7] Dow teorie: Praktické použití nejen na FOREXu | FXstreet.cz. FXstreet.cz - forex, komodity, kryptoměny, trading, zpravodajství | FXstreet.cz [online]. [cit. 2019-06-02]. Dostupné z: <https://www.fxstreet.cz/dow-teorie-prakticke-pouziti-nejen-na-forexu.html?fbclid=IwAR1FrthQo8FgemrRN0FNRnDOV4KkSPxawTPpxo83SZgNvYXf0N6-Q9BQOjw>
- [8] DUSPIVA, P. a Liběna TETŘEVOVÁ. Kapitálové trhy: pro kombinovanou formu studia. Vyd. 3., upr. a dopl. Pardubice: Univerzita Pardubice, 2006. ISBN 80-7194-896-9.
- [9] Elliottovy vlny 2019 - komplexní průvodce + bonus strategie. Admiral Markets - Obchodujte Forex, CFD, komodity CFD a více s oceňovaným online brokerem. [online]. [cit. 2019-06-02] Dostupné z: <https://admiralmarkets.cz/education/articles/forex-indicators/vysvetleni-ukazatele-elliottovy-vlny>
- [10] Finance. yahoo.com. [online]. [cit. 2019-06-15]. Dostupné z: <https://finance.yahoo.com>
- [11] HENDL, J. Kvalitativní výzkum: základní teorie, metody a aplikace. 2., aktualiz. vyd. Praha: Portál, 2008. ISBN 978-80-7367-485-4.

- [12] HENRY, E. Are Investors Influenced By How Earnings Press Releases Are Written? *Journal of Business Communication*. 2008, roč. 45, č. 4, s. 363-407. ISSN 2329-4892.
- [13] How Much Data Is Generated Per Minute? The Answer Will Blow Your Mind Away. *Digital Information World* [online]. [cit. 2019-05-12] Dostupné z: <https://www.digitalinformationworld.com/2018/06/infographics-data-never-sleeps-6.html>
- [14] How to calculate the Abnormal Return (AR)?. *ResearchGate*. Find and share research [online][cit. 2019-06-19] Dostupné z: https://www.researchgate.net/post/How_to_calculate_the_Abnormal_Return_AR
- [15] HU, M., LIU, B. Mining and Summarizing Customer Reviews. *Department of Computer Science*. [Online] 2004. [Citace: 18. 02 2019.] <https://www.cs.uic.edu/~liub/publications/kdd04-revSummary.pdf>.
- [16] Index Nasdaq - Vše co musíte vědět o US technologickém indexu. *Broker LYNX*. Investujte s výhodami [online] [cit. 2019-06-18] Dostupné z: <https://www.lynxbroker.cz/vzdelavani/index-nasdaq/>
- [17] JÍLEK, J. *Finanční trhy*. Praha: Grada, 1997. ISBN 80-7169-453-3.
- [18] JÍLEK, J. *Finanční trhy pro národohospodáře*. Praha: Vysoká škola ekonomická, 1997. ISBN 80-7079-860-2.
- [19] LIU, B.. *Sentiment analysis and opinion mining*. San Rafael: Morgan & Claypool Publishers, 2012. *Synthesis lectures on human language technologies*, 16. ISBN 9781608458844.
- [20] *Logistická regrese*. www.trilobyte.cz [online]. [cit. 2019-06-30]. Dostupné z: <http://www.trilobyte.cz/downloadfree/qcemanual/logreg.pdf>
- [21] LOUGHRAN, T., MCDONALD, B. When Is a Liability Not a Liability? *Textual Analysis, Dictionaries, and 10-Ks*. *The Journal of Finance*. 2011, roč. 66, č. 1, s. 35-65. ISSN 1540-6261.
- [22] *Obchodujeme Elliottovy vlny I. - Praxe - Finančník.cz*. *Finančník.cz - komodity, akcie, burza, forex - Finančník.cz* [online]. [cit.2019-06-02] Dostupné z: https://www.financnik.cz/clanky/praxe/elliottovy-vlny-1r316?fbclid=IwAR3crJ0NkkZ5xOqeNpOJRWR0S7GIhoDcjJ6dUQApZ2YIVfVdgt9N9Sxv_0g

- [23] PETERSON, R. L. Trading on sentiment: the power of minds over markets. Hoboken, New Jersey: Wiley, 2016. ISBN 978-1-119-12276-0.
- [24] Rambocas, M., Gama, J. Marketing Research: The Role of Sentiment Analysis, 2013, ISSN 0870-8541.
- [25] REJNUŠ, O.. Finanční trhy. 4., aktualiz. a rozš. vyd. Praha: Grada, 2014. Partners. ISBN 978-80-247-3671-6.
- [26] Reuters. Business & Financial News, U.S & International Breaking News [online]. [cit. 2019-06-15]. Dostupné z: <https://www.reuters.com/>
- [27] Úvod do regresní analýzy. www.statsoft.cz [online]. 2014 [cit. 2019-06-20]. Dostupné z:
http://www.statsoft.cz/file1/PDF/newsletter/2014_26_03_StatSoft_Uvod_do_regresni_analyzy.pdf
- [28] VALACH, J. Investiční rozhodování a dlouhodobé financování. 3., přeprac. a rozš. vyd. Praha: Ekopress, 2010. ISBN 978-80-86929-71-2.

SEZNAM PŘÍLOH

Příloha A - Výsledky logit regresního modelu [vlastní zpracování]

PŘÍLOHA

Příloha A – Výsledky logit regresního modelu

AAPL	t+1			t+3			t+5		
	Stupně volnosti	Wald. Stat.	p	Stupně volnosti	Wald. Stat.	p	Stupně volnosti	Wald. Stat.	p
Abs. člen	1	5,0321	0,02481	1	0,2646	0,6069	1	2,1177	0,1455
Počet slov ve zprávě	1	1,1759	0,1846	1	0,8975	0,3434	1	2,0862	0,1486
Poměr-Liu	1	1,1813	0,277	1	4,2065	0,0402	1	0,0581	0,8094
Abs. člen	1	2,1357	0,1436	1	2,9848	0,084	1	2,6932	0,1007
Počet slov ve zprávě	1	0,801	0,3707	1	1,4504	0,2284	1	2,7551	0,0969
Poměr-LM	1	0,0407	0,8399	1	1,0758	0,2996	1	0,0185	0,8916
Abs. člen	1	2,9154	0,0877	1	0,0541	0,816	1	0,3372	0,5614
Počet slov ve zprávě	1	0,8791	0,3484	1	0,0146	0,9037	1	0,7681	0,3807
Poměr-Henry	1	3,5066	0,0611	1	12,940	0,0003	1	2,8975	0,0887

AMZN	t+1			t+3			t+5		
	Stupně volnosti	Wald. Stat.	p	Stupně volnosti	Wald. Stat.	p	Stupně volnosti	Wald. Stat.	p
Abs. člen	1	0,0161	0,8988	1	0,0066	0,9349	1	2,1572	0,1418
Počet slov ve zprávě	1	0,1105	0,7395	1	1,8488	0,1739	1	0,0087	0,9253
Poměr-Liu	1	1,2447	0,2645	1	0,0541	0,8159	1	0,3338	0,5633
Abs. člen	1	0,0353	0,8508	1	0,2887	0,591	1	5,8531	0,0155
Počet slov ve zprávě	1	0,2399	0,6242	1	1,7996	0,1797	1	0,1332	0,7151
Poměr-LM	1	0,4547	0,5	1	1,2987	0,2544	1	3,5306	0,0602
Abs. člen	1	0,3206	0,5712	1	0,4775	0,4895	1	4,1199	0,0423
Počet slov ve zprávě	1	0,0081	0,9282	1	1,0117	0,3144	1	0,3041	0,5812
Poměr-Henry	1	1,5154	0,2183	1	0,0806	0,7763	1	0,8482	0,357

GE	t+1			t+3			t+5		
	Stupně volnosti	Wald. Stat.	p	Stupně volnosti	Wald. Stat.	p	Stupně volnosti	Wald. Stat.	p
Abs. člen	1	2,1573	0,1418	1	13,3233	0,0002	1	12,9318	0,0003
Počet slov ve zprávě	1	1,2118	0,2709	1	0,4124	0,5207	1	0,0047	0,9452
Poměr-Liu	1	0,3148	0,5747	1	0,4888	0,4844	1	0,2818	0,5954
Abs. člen	1	3,2964	0,0694	1	8,886	0,0028	1	7,3649	0,0066
Počet slov ve zprávě	1	0,4404	0,5068	1	0,0583	0,809	1	0,1157	0,7336
Poměr-LM	1	0,1035	0,7476	1	0,0536	0,8167	1	0,4725	0,4918
Abs. člen	1	0,582	0,4455	1	8,3886	0,0037	1	14,8308	0,0001
Počet slov ve zprávě	1	1,1146	0,291	1	0,1778	0,6732	1	0,2452	0,6204
Poměr-Henry	1	1,2869	0,2566	1	0,3952	0,5296	1	1,4933	0,2216

JPM	t+1			t+3			t+5		
	Stupně volnosti	Wald. Stat.	p	Stupně volnosti	Wald. Stat.	p	Stupně volnosti	Wald. Stat.	p
Abs. člen	1	1,2936	0,2553	1	0,3179	0,5728	1	0,0452	0,8315
Počet slov ve zprávě	1	2,1365	0,1438	1	4,2813	0,0385	1	0,2155	0,6424
Poměr-Liu	1	0,7914	0,3736	1	0,0054	0,9409	1	0,1155	0,7339
Abs. člen	1	0,4845	0,4863	1	1,9125	0,1666	1	0,0202	0,8868
Počet slov ve zprávě	1	1,5184	0,2178	1	6,305	0,012	1	1,0879	0,2969
Poměr-LM	1	0,1894	0,6633	1	0,0849	0,7706	1	0,6832	0,4084
Abs. člen	1	0,1149	0,7346	1	0,5238	0,4691	1	0,2846	0,5936
Počet slov ve zprávě	1	2,0963	0,1476	1	5,4505	0,0195	1	0,2853	0,5932
Poměr-	1	1,0669	0,3016	1	4,2626	0,0369	1	1,2373	0,2659

Henry			
-------	--	--	--

MSFT	t+1			t+3			t+5		
	Stupně volnosti	Wald. Stat.	p	Stupně volnosti	Wald. Stat.	p	Stupně volnosti	Wald. Stat.	p
Abs. člen	1	1,0265	0,3109	1	3,4454	0,0634	1	0,9929	0,319
Počet slov ve zprávě	1	0,7391	0,3899	1	0,483	0,487	1	0,3268	0,5675
Poměr-Liu	1	0,6386	0,4241	1	0,3005	0,5835	1	0,0366	0,8482
Abs. člen	1	0,6094	0,4349	1	2,2561	0,133	1	0,0915	0,7622
Počet slov ve zprávě	1	0,6085	0,4353	1	0,3198	0,5716	1	0,686	0,4074
Poměr-LM	1	0,4298	0,512	1	0,0498	0,8232	1	0,4487	0,5029
Abs. člen	1	0,4065	0,5237	1	1,269	0,2599	1	0,1906	0,6623
Počet slov ve zprávě	1	1,9096	0,1669	1	0,2197	0,6392	1	0,2687	0,2814
Poměr-Henry	1	0,2749	0,6	1	2,4654	0,1163	1	1,1601	0,2814

PFE	t+1			t+3			t+5		
	Stupně volnosti	Wald. Stat.	p	Stupně volnosti	Wald. Stat.	p	Stupně volnosti	Wald. Stat.	p
Abs. člen	1	0,6001	0,4385	1	0,2163	0,6418	1	0,5806	0,446
Počet slov ve zprávě	1	2,0241	0,1548	1	0,3494	0,5544	1	0,7871	0,3749
Poměr-Liu	1	0,7336	0,3918	1	1,3971	0,2371	1	0,3256	0,5682
Abs. člen	1	0,0985	0,7535	1	1,4124	0,2346	1	1,7732	0,1829
Počet slov ve zprávě	1	1,6806	0,1948	1	0,1007	0,7509	1	0,8856	0,3466
Poměr-LM	1	0,0276	0,8679	1	1,7443	0,1865	1	0,1052	0,7456
Abs. člen	1	0,0167	0,8971	1	0,0145	0,904	1	0,2918	0,589
Počet slov ve zprávě	1	0,6009	0,4382	1	0,0304	0,8613	1	0,0223	0,8812

zprávě									
Poměr-Henry	1	0,359	0,549	1	0,055	0,8144	1	0,625	0,4291
	t+1			t+3			t+5		
T	Stupně volnosti	Wald. Stat.	p	Stupně volnosti	Wald. Stat.	p	Stupně volnosti	Wald. Stat.	p
Abs. člen	1	1,0265	0,3109	1	3,4454	0,0634	1	0,9929	0,319
Počet slov ve zprávě	1	0,6386	0,4241	1	0,483	0,487	1	0,3268	0,5675
Poměr-Liu	1	0,7391	0,3899	1	0,3005	0,5835	1	0,0366	0,8482
Abs. člen	1	0,6094	0,4349	1	2,2561	0,133	1	0,0915	0,7622
Počet slov ve zprávě	1	0,6085	0,4353	1	0,3196	0,5716	1	0,686	0,4074
Poměr-LM	1	0,4298	0,512	1	0,0498	0,8232	1	0,4487	0,5029
Abs. člen	1	0,4065	0,5237	1	1,269	0,2599	1	0,1906	0,6623
Počet slov ve zprávě	1	1,9096	0,1669	1	0,2197	0,6392	1	0,2687	0,6041
Poměr-Henry	1	0,2749	0,6	1	2,4654	0,1163	1	1,1601	0,2814

	t+1			t+3			t+5		
WFC	Stupně volnosti	Wald. Stat.	p	Stupně volnosti	Wald. Stat.	p	Stupně volnosti	Wald. Stat.	p
Abs. člen	1	0,1473	0,701	1	0,6242	0,4294	1	0,2438	0,6214
Počet slov ve zprávě	1	2,0228	0,1549	1	2,6247	0,1052	1	6,2484	0,0124
Poměr-Liu	1	0,8407	0,3591	1	2,8464	0,0915	1	4,4675	0,0345
Abs. člen	1	0,0008	0,977	1	0,0177	0,8941	1	0,7349	0,3912
Počet slov ve zprávě	1	2,036	0,1536	1	2,5416	0,1108	1	5,2295	0,0222
Poměr-LM	1	0,021	0,8846	1	0,3061	0,58	1	2,3099	0,1285
Abs. člen	1	0,0827	0,7735	1	0,093	0,7603	1	0,6259	0,4288
Počet slov ve zprávě	1	1,0058	0,3158	1	1,7646	0,184	1	1,5235	0,217

Poměr-Henry	1	0,0022	0,9624	1	2,7401	0,0978	1	2,3413	0,1259
-------------	---	--------	--------	---	--------	--------	---	--------	--------

WMT	t+1			t+3			t+5		
	Stupně volnosti	Wald. Stat.	p	Stupně volnosti	Wald. Stat.	p	Stupně volnosti	Wald. Stat.	p
Abs. člen	1	0,2776	0,5982	1	1,2996	0,2542	1	4,2628	0,0389
Počet slov ve zprávě	1	0,007	0,9332	1	0,0777	0,7804	1	1,3948	0,2375
Poměr-Liu	1	0,7705	0,38	1	2,7819	0,0953	1	4,3444	0,0371
Abs. člen	1	0,058	0,8096	1	1,4622	0,2265	1	5,2714	0,0216
Počet slov ve zprávě	1	0,0712	0,7895	1	0,994	0,3187	1	4,0949	0,043
Poměr-LM	1	0,7535	0,3853	1	2,9905	0,0837	1	0,8649	0,3523
Abs. člen	1	1,0119	0,3143	1	1,7414	0,1869	1	2,6916	0,1008
Počet slov ve zprávě	1	1,3154	0,2514	1	0,1504	0,6981	1	0,497	0,4808
Poměr-Henry	1	0,2811	0,5959	1	1,6238	0,2025	1	2,893	0,0889

XOM	t+1			t+3			t+5		
	Stupně volnosti	Wald. Stat.	p	Stupně volnosti	Wald. Stat.	p	Stupně volnosti	Wald. Stat.	p
Abs. člen	1	4,3045	0,038	1	0,173	0,6773	1	0,6984	0,4033
Počet slov ve zprávě	1	0,3164	0,5737	1	0,3264	0,5677	1	0,1638	0,6856
Poměr-Liu	1	1,2251	0,2683	1	0,9613	0,3268	1	0,0078	0,9293
Abs. člen	1	1,7283	0,1886	1	0,0761	0,7825	1	0,4046	0,5246
Počet slov ve zprávě	1	0,2107	0,6462	1	1,4948	0,2214	1	0,5016	0,4787
Poměr-LM	1	0,5518	0,4575	1	1,9735	0,16	1	0,8822	0,3475
Abs. člen	1	4,1823	0,0408	1	0,5191	0,4712	1	0,1039	0,747
Počet slov ve	1	0,6276	0,4282	1	0,1557	0,6932	1	0,3377	0,5611

zpráve									
Poměr- Henry	1	3,9521	0,0468	1	1,3434	0,2464	1	3,3676	0,0664