

Univerzita Pardubice

Fakulta ekonomicko-správní

Predikce akciových výnosů pomocí analýzy sentimentu sociálních médií

Bc. Patrik Charamza

Diplomová práce

2019

Zadávací list

Zadávací list

Prohlašuji:

Tuto práci jsem vypracoval samostatně. Veškeré literární prameny a informace, které jsem v práci využil, jsou uvedeny v seznamu použité literatury.

Byl jsem seznámen 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.

PODĚKOVÁNÍ:

Tímto bych rád poděkoval především svému vedoucímu práce, doc. Ing. Petru Hájkovi, Ph.D. nejen za odbornou asistenci, materiály a užitečné rady, ale také za vstřícný a flexibilní přístup. Dále chci projevit svůj vděk i pracovnícům studijního oddělení fakulty ekonomicko – správní, za poskytování perfektního administrativní zázemí.

Děkuji konečně také všem mým blízkým za podporu a trpělivost, kterou v mi době mé tvorby práce projevili.

ANOTACE

Tato diplomová práce je zaměřena na průzkum sociálních médií pomocí analýzy sentimentu. Cílem je zjistit, zda lze využít získaná data k predikci vývoje akciových kurzů. Je zde zkoumán vztah mezi sentimentem vyjadřovaným v internetových diskuzích a reálným vývojem ceny akcií. Výsledek je porovnán s oblíbenými indikátory technické analýzy. Data k osmi vybraným akciovým titulům byla získána ze sociální sítě Yahoo Finance.

KLÍČOVÁ SLOVA

Analýza sentimentu, sentiment, technická analýza, sociální média, predikce, akciové výnosy

ANNOTATION

This thesis is an exploration of social media via sentiment analysis. The main objective is to find out if it is possible to make predictions of development of stock exchange rates based on data from online discussions. The relation between sentiment expressed in comments and real development of stock prices is examined. The result is compared with traditional indicators of technical analysis. Data for eight stocks was obtained from social media Yahoo Finance.

KEYWORDS

Sentiment, Sentiment Analysis, Technical Analysis, Social Media, Prediction, Stock Yields

OBSAH

ÚVOD.....	12
1. Analýzy finančních trhů	14
1.1. Hospodářské cykly	15
1.2. Fundamentální analýza.....	17
1.2.1. Globální fundamentální analýza.....	17
1.2.2. Odvětvová fundamentální analýza	18
1.2.3. Firemní fundamentální analýza	18
1.3. Technická analýza	19
1.3.1. Dowova teorie	20
1.3.2. Elliottovy vlny.....	21
1.3.3. Indikátory technické analýzy	23
1.4. Psychologická analýza	25
1.4.1. Příslušnost k davu	26
1.4.2. Vliv emocí.....	27
1.4.3. Názor	28
1.4.4. Sentiment	28
1.5. Analýza textu	30
1.5.1. Slovníky	34
1.5.2. Specifika debat a komentářů	35
2. Analýza sentimentu sociálních médií	36
2.1. Příprava dat	36
2.1.1. Sociální média.....	36
2.1.2. Sběr a čištění dat	38
2.2. Měření sentimentu.....	39
2.2.1. Analýza komentářů	41
2.2.2. Vývoj sentimentu	44
2.3. Predikce změn ceny akcií na základě analýzy sentimentu.....	52
2.3.1. Popis proměnných.....	52
2.3.2. Faktorová analýza	54
2.3.3. Regresní analýza	56

2.3.4. Vliv technických indikátorů a sentimentu.....	59
ZÁVĚR	64
POUŽITÁ LITERATURA	65
SEZNAM PŘÍLOH.....	69

SEZNAM TABULEK

Tabulka 1: Příklad analýzy sentimentu na vybraném komentáři	40
Tabulka 2: Nejčastější slova v diskuzi AAPL, AMZN.....	41
Tabulka 3: Nejčastější slova v diskuzi BABA, BAC	42
Tabulka 4: Nejčastější slova v diskuzi CSCO, FB.....	43
Tabulka 5: Nejčastější slova v diskuzi T, WMT.....	44
Tabulka 6: Korelace proměnných vstupujících do regrese	54
Tabulka 7: Korelace proměnných se zastoupením faktorem	56
Tabulka 8: Výsledky regresní analýzy	59

SEZNAM ILUSTRACÍ

Obrázek 1: Schéma finančních trhů.....	15
Obrázek 2: Vývoj HDP v ČR v letech 2005–2015.....	16
Obrázek 3: Dowova teorie	21
Obrázek 4: Elliottovy vlny	22
Obrázek 5: Sociální sítě dle počtu uživatelů	37
Obrázek 6: Vývoj sentimentu Apple Inc.	45
Obrázek 7: Vývoj sentimentu Amazon.com, Inc.	46
Obrázek 8: Vývoj sentimentu Alibaba Group Holding Limited.	47
Obrázek 9: Vývoj sentimentu Alibaba Group Holding Limited	48
Obrázek 10: Vývoj sentimentu Cisco Systems, Inc.	49
Obrázek 11: Vývoj sentimentu Facebook, Inc.	50
Obrázek 12: Vývoj sentimentu AT&T Inc.	51
Obrázek 13: Vývoj sentimentu Walmart Inc.	52
Obrázek 14: Metoda nejmenších čtverců	57
Obrázek 15: Regresní analýza v software STATISTICA	58
Obrázek 16: Předpovědi vs rezidua, software STATISTICA	61
Obrázek 17: Normální p – graf reziduí, software STATISTICA	62
Obrázek 18: Histogram reziduí, software STATISTICA	63

SEZNAM ZKRATEK A ZNAČEK

10-Ks	Zprávy o výkonu společností používané ve Spojených státech amerických
AAPL	Akciový titul Apple Inc.
AMZN	Akciový titul Amazon.com, Inc.
BABA	Akciový titul Alibaba Group Holding Limited
BAC	Akciový titul Bank of America Corporation
CCI	Index komoditních kanálů (Commodity Channel Index)
CSCO	Akciový titul Cisco Systems, Inc
CSR	Sekvenční třídící pravidlo (Class Sequential Rule)
CVX	Akciový titul Chevron Corporation
EMA	Exponenciální klouzavý průměr (Exponential moving average)
FB	Akciový titul Facebook, Inc.
H4N	Označení Harvardského slovníku Harvard-IV-4 TagNeg
HDP	Hrubý domácí produkt
IE	Extrakce informací (Information Extraction)
INTC	Akciový titul Intel Corporation
IR	Vyhledávání a získávání informací (Search and Information Retrieval)
JNJ	Akciový titul Johnson & Johnson
JPM	Akciový titul JPMorgan Chase & Co.
KO	Akciový titul The Coca-Cola Company
MACD	Klouzavý průměr konvergence divergence (Moving Average Convergence Divergence)
MSFT	Akciový titul Microsoft Corporation
NLP	Zpracování přirozeného jazyka (Natural Language Processing)
P/BV	Poměr ceny a účetní hodnoty společnosti (Price/Book Value ratio)
P/E	Poměr mezi tržní cenou akcie a čistým ziskem na akcii (Price-to-Earnings ratio)
P/S	Poměr tržní ceny akcie k tržbám společnosti (Price/Sales ratio)
PFE	Akciový titul Pfizer Inc
PG	Akciový titul The Procter & Gamble Company
POS	Metoda slovních druhů (Part of Speech)
RSI	Index relativní síly (Relative Strength Index)
SMA	Jednoduchý klouzavý průměr (Simple moving average)
SO	Stochastický oscilátor

T	Akciový titul AT&T Inc.
V	Akciový titul Visa Inc.
VZ	Akciový titul Verizon Communications Inc.
WFC	Akciový titul Wells Fargo & Company
WMT	Akciový titul Walmart Inc.
XOM	Akciový titul Exxon Mobil Corporation

ÚVOD

Tato práce zkoumá vývoj ceny vybraných akciových titulů na finančních trzích dvěma způsoby. V první řadě tradičním přístupem technické analýzy, dále pak aplikací modernějšího přístupu – analýzy sentimentu sociálních médií. Finanční analytici se snaží už mnoho let najít recept na zaručený zisk, pozorováním vývoje fundamentálních makroekonomických veličin, zkoumáním psychologie investorů, ale také různými výpočty, indikátory a studiem grafických formací. Dnes však mají k dispozici zdroje informací, které dříve neexistovaly. S rozvojem komunikačních technologií a rozmachem sociálních sítí vznikl virtuální prostor, ve kterém se šíří informace v obrovském množství a rychleji než kdy dříve. Každá z těchto informací může ve finančním světě být klíčem k úspěchu, proto se analýzy sentimentu a data mining obecně stávají terčem pozornosti jak individuálních investorů, tak velkých korporací. Dolování užitečných dat z dokumentů či webových stránek není ale pochopitelně metodou aplikovatelnou pouze ve finanční sféře. Velké uplatnění dnes nalézají také v oblasti zboží a služeb, marketingu, turismu, politiky a v neposlední řadě také při prevenci kriminality.

Tato práce se v první části zaměřuje na shrnutí teoretický východisek, základních principů fungování finančních trhů a představení tři základních větví metod zkoumání finančních trhů. Fundamentální přístupy nejsou pro praktickou část práce zásadní, větší pozornost je věnována metodám technické a psychologické analýzy. Shrnutí jsou nejen historické myšlenky, které byly za zrodem technické analýzy, ale také konkrétní metody využívané v dnešní době. V pojednání o psychologickém aspektu obchodování jsou definovány klíčové pojmy jako názor, emoce či sentiment, se kterými pracuje druhá část práce. Také je zde přiblíženo moderní pojetí analýzy sentimentu na různých úrovních a užití tohoto univerzálního nástroje v odvětví financí a investic.

Ve druhé části práce je výstup z praktické aplikace analýzy sentimentu na sociální síti Yahoo Finance. Nejprve jsou objasněny důvody výběru konkrétní sociální sítě i jednotlivých akciových titulů. Poté byla z uživatelských diskuzí k těmto titulům shromážděna data ve formě komentářů. Sběr dat probíhal u osmi akciových titulů v době od 1.11.2018 do 31.12.2018. Po očištění a úpravě dat mezi nimi byly hledány výrazy nesoucí pozitivní či negativní sentiment. Pro tento krok byly použity tři slovníky od různých autorů: Elaine Henry, Bing Liu a dvojice Tima Loughrana a Billa McDonalda. Vypočtený výsledný sentiment je potom porovnán s vývojem ceny akciových titulů a s tradičními ukazateli technické analýzy. Pro posouzení možnosti využití výstupu analýzy sentimentu k predikci vývoje ceny akcií je použita regresní

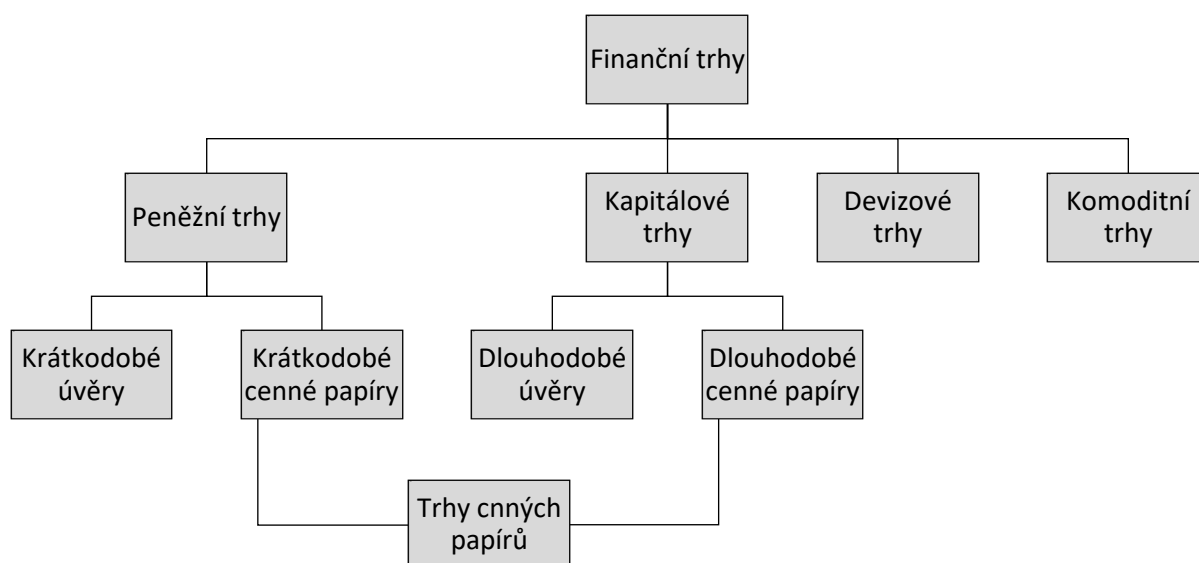
analýza. Tou je v závěru odhaleno, zda byl ve sledovaném období vývoj ceny závislý na zjištěném sentimentu.

Cílem práce je tedy představit současné přístupy k analýze sentimentu a upozornit na některé problémové oblasti tohoto přístupu. Dále je cílem užitím analýzy sentimentu zjistit vývoj sentimentu u konkrétních akciových titulů v daném období a porovnat jej s vývojem ceny akcií těchto titulů. Posledním cílem je užitím regresní analýzy zjistit, zda mají výsledná data potenciál pro predikci vývoje ceny akcií. Získané poznatky a dosažení stanovených cílů je zhodnoceno v závěru práce.

1. Analýzy finančních trhů

Tržní prostředí v moderních ekonomikách nabízí dnes řadu instrumentů a investičních nástrojů, mezi kterými jednotliví investoři hledají možnosti, jak dosáhnout zisku. Akcie patří mezi tradičnější a nejrozšířenější finanční instrumenty. Je to cenný papír, představující podíl na vlastnictví akciové společnosti [16]. Akcií se rozlišuje více druhů a z jejich držení přísluší vlastníkovu určitá práva. Význam podnikových akcií spočívá v první řadě v tom, že jejich emitováním a následným prodejem na primárním trhu si akciové společnosti opatřují potřebné peněžní prostředky, jež v budoucnu nemusí vrátit [34]. Akcionáři se také mohou podílet na řízení společnosti a na jejím zisku, případně likvidačním zůstatku [16]. Tím ale možnosti investorů nekončí, akcie jsou cenný papír s vysokou volatilitou, motivem nákupu akcií tedy bývá též snaha investorů dosáhnout vedle dividend rovněž výnosů kapitálových, vyplývajících z jimi předpokládaných změn jejich tržních cen [34].

Tržní cena akcie vychází z mnoha faktorů. Podle [20] je hodnota akcií dána tím, jak kapitálový trh odhaduje hodnotu celkové sumy dividend, které by podnik mohl teoreticky vyplatit. Kromě dividendové politiky se ale obvykle posuzuje i finanční strukturu firmy, schopnost generovat zisk a plnit další hospodářské cíle. Posouzení těchto stránek je však vždy do jisté míry subjektivní a tím lze přejít k samotnému principu fungování finančních trhů. Tržní cena akcií se neustále mění v závislosti na tom, jak subjekty na trhu nahlíží na již zmíněné faktory a také celou řadu dalších relevantních informací a jelikož se tyto názory i informace různí, různí se také představy o správné tržní ceně. V důsledku toho lze považovat určitou akcii za podhodnocenou, což motivuje subjekty akcii nakupovat, anebo nadhodnocenou, což motivuje subjekty akcii prodat. Podle [34] představuje právě tato spekulace významný faktor integrace segmentů finančního trhu. Podle jedné teorie má vývoj cen akcií charakter náhodné procházky (random walk), kdy cena akcií je výsledkem mnoha náhodných událostí a nelze ji předpovídat [16]. V jeden okamžik se každopádně mohou střetnout dva subjekty, jeden se zájmem akcii prodat, druhý se zájmem akcii nakoupit a akciový trh je rázem na světě. Jeho pozici ve struktuře finančního trhu ilustruje následující grafika (obrázek 1).



Obrázek 1: Schéma finančních trhů

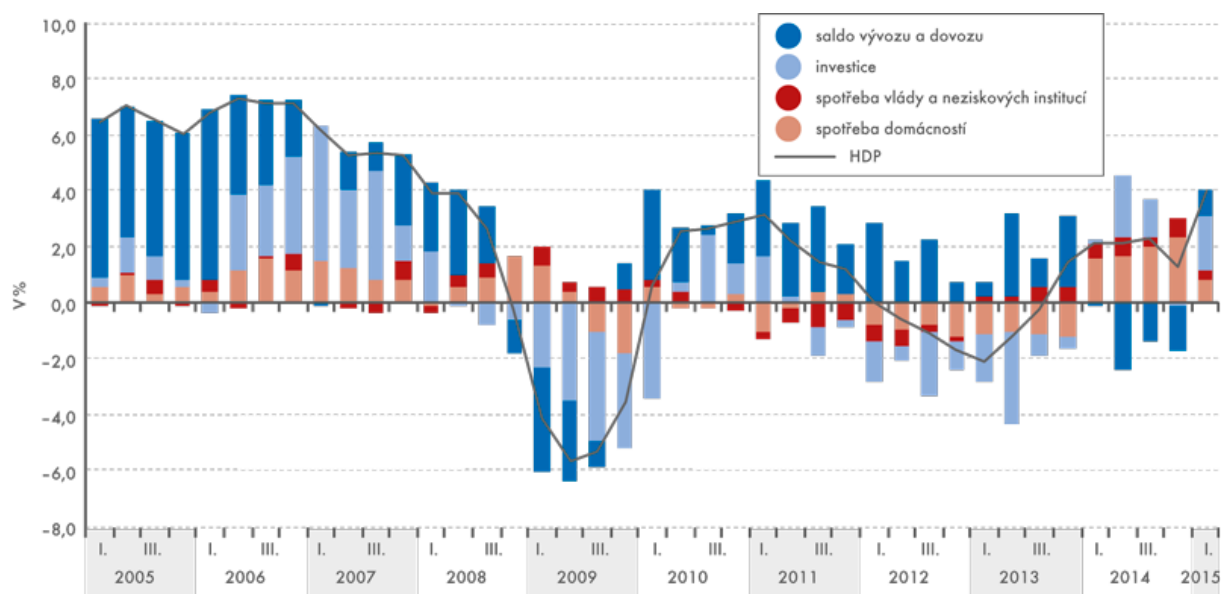
Zdroj: Vlastní zpracování dle [33]

1.1. Hospodářské cykly

Analytické metody napříč finančním trhem při snaze o predikci zkoumají historická data a hledají určité zákonitosti. Každý existující finanční systém je vždy nedílnou součástí některého systému ekonomického [34]. Jako celek funguje podle určitých „vzorců“ a některé zdánlivě náhodné události se opakují pravidelně. Tyto pravidelnosti v ekonomickém prostředí mají formovat jev, kterému se říká ekonomický či hospodářský cyklus.

Faktem ale zůstává, že ačkoliv historie potvrzuje jejich výskyt, neexistují žádné důkazy potvrzující konstantní pravidelnost těchto cyklů. Podle [20] lze právě z tohoto důvodu hovořit pouze o dojmu hospodářských cyklů. Absence pravidelnosti znamená, že ve skutečnosti nejde o cyklus, ale spíše o výkyvy konstantně se měnícího prostředí. Joseph Kitchin v roce 1923 odhalil krátkodobý, tří – až pětiletý obchodní cyklus založený na přizpůsobování zásob. Clement Juglar zaznamenal v roce 1862 střednědobý, sedmi – až jedenáctiletý cyklus související s kapitálovými investicemi. Simon Kuznets v roce 1930 poukázal na 12-25leté cykly související s výstavbou a demografickými faktory. Brian Berry poukázal na přítomnost generaci trvajících, 25 - 35letého cyklu souvisejícího s vývojem investic do infrastruktury. Nikolaj Kondratěv zase v roce 1926 zveřejnil svůj slavný článek o dlouhodobých vlnách v cenách komodit trvajících 40 až 60 let. A konečně v roce 1991 prezentovali američtí ekonomové William Strauss a Neil Howe přesvědčivý důkaz o existenci mezigeneračního, 85–

99letého krizového cyklu [39]. Ekonomické nebo také hospodářské cykly mají stejnou dynamiku jako cykly na finančních trzích. Je jasné, že výnosy cenných papírů (zejména akcií) jsou do značné míry určeny vývojem hrubého domácího produktu – už jen z toho důvodu, že jeho součástí jsou také zisky podniků [20]. Investoři ale pochopitelně nechtějí čekat desítky let na dovršení jednoho ze zmíněných cyklů, proto zkoumají veškeré cyklické formace v různých obdobích. Stejný vývoj, jako mají dlouhodobé cykly, lze vysledovat i na měsíčních, týdenních, denních nebo minutových grafech (obrázek 2).



Obrázek 2: Vývoj HDP v ČR v letech 2005–2015

Zdroj: [36]

V roce 2015 byl na základě dat od Českého statistického úřadu zkonstruován graf (obrázek 2), na kterém můžeme vidět cyklický vývoj HDP v České republice, v letech 2005 až 2015. Růst je střídán poklesem, pokles poté růstem, následně se opakuje pokles. Rozdíl mezi lety 2014 a 2015 byl komentován následovně: „Příznivý hospodářský vývoj se projevil napříč celou tuzemskou ekonomikou. Meziročně rostla nejen většina odvětví průmyslu, ale také podnikové služby napojené na rozvoj ostatních odvětví. Zpracovatelský průmysl vzrostl meziročně o 6,3 % a v porovnání s posledním čtvrtletím loňského roku se zvýšil o 2,6 %. Na tento růst měl největší vliv automobilový průmysl, ale také ostatní tradiční odvětví zpracovatelského průmyslu, jako například chemický průmysl, výroba pryže a plastů či výroba elektrických zařízení a výroba kovů a kovových konstrukcí“ [36].

Dále se v analýze hovoří o návratu k předkrizovému vývoji, což opět lze označit za určitý cyklus. Důležitá je ale i zmínka o růstu v různých odvětvích. Zpětně lze vidět, ve který moment

by byl vhodný okamžik k investici a kterým odvětvím se dařilo nejvíce. Jestli ale bylo v tom daném momentu v historii možné tento budoucí vývoj spolehlivě vyčíst z vývoje ostatních odvětví a celé ekonomiky, je otázka.

Pro úspěšné obchodování je tedy stěžejní správný úsudek a také správné načasování. Při snaze o co nejúspěšnější prognózu investoři využívají zaprvé metody založené na různých informacích o společnosti včetně poměru P/E, výročních zprávách, dění v podnicích a makroekonomickém vývoji. Druhá metoda je založena na vývoji cen a objemu obchodování s akciami v minulosti. Třetí je založena na odhadu budoucího jednání lidí [16]. V průběhu staletí byly teoretiky i praktiky vytvořeny tyto tři zcela rozdílné analytické přístupy. Jedná se o [39]:

- fundamentální analýzu,
- technickou analýzu,
- psychologickou analýzu.

1.2. Fundamentální analýza

Fundamentální analýza je nejkompexnějším a přístupem, který se pokouší vysvětlit pohyb akciových kurzů. Často v ní analytici porovnávají vnitřní hodnotu akcie s tržní hodnotou, vnitřní hodnotu však definují poměrně obecně a mlhavě [16]. Dále se zabývá detailním zkoumáním základních a podstatných ekonomických, politických, sociálních, geografických, demografických a jiných faktorů a událostí, které determinují vývoj akciových kurzů [39]. Největším problémem fundamentální analýzy je obtížnost získávání, a hlavně věrohodnost potřebných dat. Jsou-li finanční údaje zkreslené, je všechno špatně [16].

Zpravidla se provádí na třech úrovních, které jsou popsány v následujících odstavcích.

1.2.1. Globální fundamentální analýza

Na této úrovni hodnotí fundamentální analýza vliv národních a nadnárodních ekonomických ukazatelů a veličin na cenu akcie. Podle toho, s jakým zpožděním ukazatele reprezentují skutečný průběh ekonomických cyklů, je lze rozdělit na předbíhající, pokrývající a zpožděné [33].

Sledován bývá růst HDP, podle [20] ale velikost soukromé spotřeby v rámci HDP neprozrazuje nic o tom, jakým způsobem lidé utrácejí své peníze, a proto nelze mechanicky porovnávat HDP mezi různými zeměmi nebo obdobími. Faktory, které předbíhají trendu, jsou například vývoj peněžní nabídky, indexy spotřebitelských očekávání, změny v cenách materiálu, nová stavební

povolení, objednávky strojního zařízení a zásob, počty poskytovaných úvěrů a jiné [33]. I vývoj akciových kurzů je však sám o sobě ekonomickým ukazatelem, a to jedním z těch nejčilejších. Pouze peněžní nabídka reaguje ještě dříve než akciové kurzy, ta je tedy velmi oblíbeným ukazatelem. Jinak je ale použití ostatních faktorů k predikci akciových kurzů značně komplikované a omezené. Jako pokrývající faktor bývá uváděno zejména HDP. Zpoždujícími se faktory jsou například míra nezaměstnanosti či inflace [33].

1.2.2. Odvětvová fundamentální analýza

Odvětvová analýza se provádí z toho důvodu, že jednotlivá odvětví nejsou stejně citlivé na celkový vývoj ekonomiky. Existuje u nich různá míra zisku, jsou různě regulovány státem a existují u nich mnohdy značně rozdílné perspektivy jejich dalšího rozvoje. Analýza se zaměřuje v první řadě na identifikaci charakteristických znaků jednotlivých odvětví či oborů [39].

Silný vliv na vývoj akcie a její tržní volatilitu a současnou a budoucí ziskovost má charakter konkurence neboli tržní struktura v odvětví působnosti firmy. Čím ostřejší konkurence v sektoru, tím vyšší nestabilita zisků i vyšší pravděpodobná volatilita akcie. Výzvou sektorové (odvětvové) fundamentální analýzy je nalézt atraktivní, nadprůměrně expanzivní odvětví s vysokou dynamikou ziskovosti firem, jejich vysokou perspektivou do budoucna [30]. Správné posouzení toho, v jaké vývojové fázi se odvětví nachází, je zcela zásadní pro úspěšnost investice. Existují dlouhodobě růstová a úspěšná odvětví, dále stagnující odvětví a odvětví v útlumu, kde se neočekává v dohledné době žádná velká inovace a zlepšení, a také mladá a technologicky nevyčerpaná odvětví, kde se předpokládá výrazný růst díky inovacím [39].

Významným kritériem je citlivost odvětví na hospodářský cyklus, zde se rozlišují [34]:

- Odvětví cyklická, která kopírují hospodářský cyklus, prosperují v období expanze a mají problémy v období recese. Důvodem je, že zákazník v těchto odvětvích může nákup snadno odložit na později, až bude ekonomická situace příznivější [33].
- Odvětví neutrální, která nebývají hospodářskými cykly příliš ovlivněna, neboť produkty z tohoto odvětví zákazník nezbytně potřebuje a nemůže jejich nákup odložit [34].
- Odvětví anticyklická, kterým se daří v období recese a mají problémy v období expanze.

1.2.3. Firemní fundamentální analýza

Nazývána také analýzou jednotlivých titulů, je procesem posuzování nadhodnocenosti nebo podhodnocenosti ceny akcií jedné konkrétní společnosti. Na této úrovni se investor snaží

odhadnout již konkrétní číslo, které představuje „správnou“ cenu akcie [39]. Tedy vypočítat tzv. vnitřní hodnotu. Na základě jejího porovnání s tržní cenou pak může investor činit své investiční rozhodnutí. K její určení opět existuje velké množství postupů a modelů, jsou to:

- Dividendové diskontní modely, kde se vnitřní hodnota určuje na základě diskontování dividend [16]. Jsou jednou z nejpropracovanějších metod, která je založena na předpokladu, že vnitřní hodnota akcie je dána současnou hodnotou veškerých budoucích příjmů z této akcie [39].
- Ziskové modely, které počítají s tím, že společnost bude každý rok vyplácet dividendy ve stejném poměru k zisku a tyto se navíc reinvestují [16], pracují s veličinou čistého zisku, kterou zpravidla různým způsobem rozkládají nebo upravují. Respektují časovou hodnotu peněz a jednoduchými matematickými úpravami lze ze ziskových modelů vyjádřit různé druhy oblíbených ukazatelů kapitálového trhu, jakými jsou P/E (price/earnings) ratio, P/BV (price/book value) ratio a P/S (price/sales) ratio [39].
- Bilanční modely nerespektují časovou hodnotu peněz a vychází ze zveřejněných účetních výkazů akciové společnosti. Lze použít model účetní hodnoty, substanční hodnoty, likvidační hodnoty, reprodukční hodnoty a substituční hodnoty [39].
- Nevýhodou diskontních a ziskových modelů bývá fakt, že část zisku, která zůstává ve společnosti zadržena, nebývá u těchto modelů zohledněná. Proto vznikly Cash flow modely využívající hotovostní toky [18]. Cash flow modely zohledňují i volné peněžní prostředky, které zůstávají na úrovni společnosti po úhradě obligatorních závazků a výplatě dividend [39].

1.3. Technická analýza

Její účelem je, stejně jako u fundamentální analýzy, predikce vývoje kurzů analyzovaných titulů a určení nejvhodnějšího okamžiku k realizaci obchodů. Na rozdíl od fundamentální analýzy však technická analýza představuje studium samotného trhu a nikoliv faktorů, které ovlivňují trh [16]. Vychází z historie a veřejně dostupných informací o burzovních indexech, objemech realizovaných obchodů a podobně. A jelikož se historie kurzových změn v průběhu času neustále opakuje, snaží se pomocí časových řad identifikovat vývojové trendy, z nichž potom vyvozují budoucí vývoj kurzů akcií [33].

1.3.1. Dowova teorie

Technická analýza je nejstarším přístupem k analyzování vývoje kurzů, její základní principy, teze a postupy teoreticky shrnul a rozpracoval nejprve Charles H. Dow (1851-1902) a poté jeho následovníci Wiliam P. Hamilton a Robert Rhea před více než 100 lety [39]. Tato teorie, přestože je již v některých částech překonána, je až doposud nejznámější a všeobecně nejuznávanější teorií tohoto druhu. Dowova teorie vychází z předpokladu, že se vývoj kurzů většiny akcií pohybuje stejným směrem, jakým se vyvíjí celý akciový trh [28]. Tento základní předpoklad je podložen následujícími poznatky:

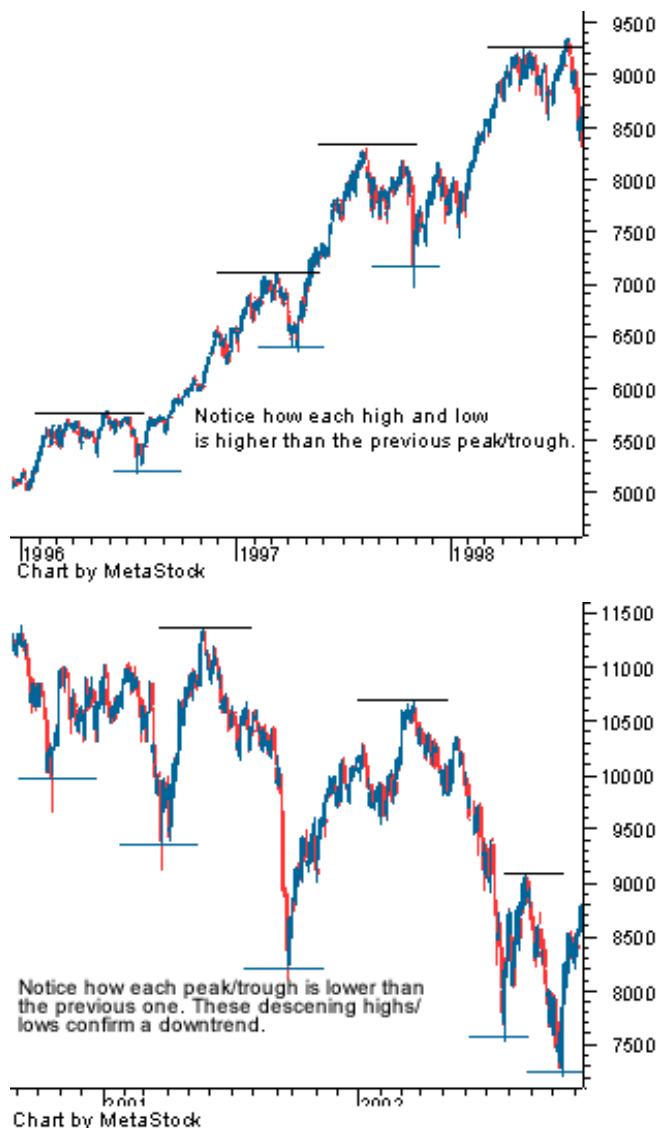
- Trh má tři trendy, tento princip vychází z lidské podstaty reagovat za stejných okolností podobně. Primární trh má tři fáze (akumulace, participace, distribuce.) Objem obchodování je důležitý sekundární faktor, který potvrzuje cenové trendy [16].
- Dění na akciových trzích se odráží ve vývoji akciových indexů „zobrazujících“ změny v chování účastníků akciového trhu [33]. Změny tržních cen jsou důsledkem agregátního pohledu všech účastníků trhu, tržní cena tedy odráží veškeré dostupné informace [16].
- Budoucí vývoj akciových burzovních indexů lze odvodit z jejich minulého vývoje, a to konkrétně z vývoje jejich uzavíracích denních hodnot [33]. Trendy existují, dokud definitivní signály neprokáží, že skončily [16].

Jakožto první ucelená teorie obsahuje určité nedokonalosti, přesto dlouhodobé výzkumy ukazují, že signály založené na Dowově teorii, jsou převážně úspěšné. Teorii bývají vyčítány následující nedostatky [33]:

- Platí pouze pro primární trend, zatímco zisky lze realizovat i na střednědobých či krátkodobých.
- Signály k nákupům a prodejům jsou indikovány příliš pozdě.
- Často se stává, že signály nejsou jednoznačné.
- Teorie nedefinuje ani dobu trvání, ani „velikost“ jednotlivých trendů.
- Zabývá se akciovým trhem jako celkem, a tudíž není vhodná pro posuzování jednotlivých titulů.

Dowova teorie bývá dokládána grafickým modelem, který je znázorněn níže (obrázek 3). Je z něho patrné, že směr trhu není dán přímoú linií, ale sérií lokálních minim a maxim neboli, primárními, sekundárními a terciálními trendy [28].

Mezi liniemi v grafu je vždy patrný sekundární trend, který má opačný směr než trend primární. Důležité však je, že ve stoupajícím primárním trendu má každý ze sekundárních trendů své linie maxima a minima výše než sekundární trend předchozí. U klesajícího primárního trendu je tomu pochopitelně naopak.



Obrázek 3: Dowova teorie

Zdroj: [28]

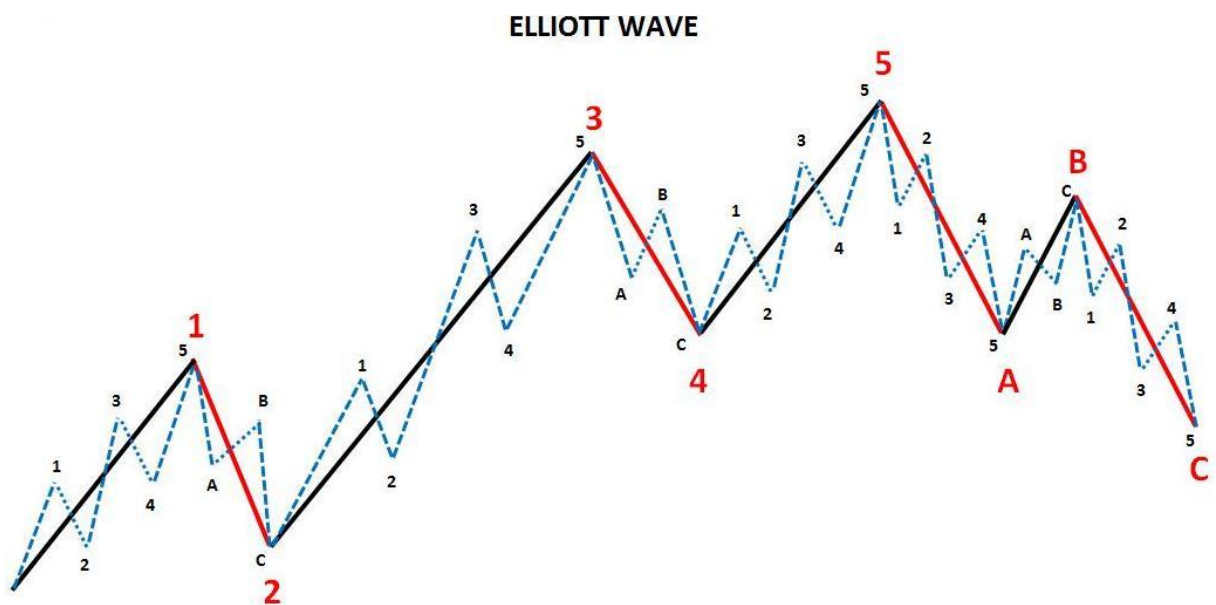
1.3.2. Elliottovy vlny

Teorie Elliottových vln se zabývá periodickým opakováním určitých „vln“ v trendu, podobně jako je popsán hospodářský cyklus v kapitole 1.1. Vychází z periodického opakování nejen

ekonomické situace, ale i přírodních jevů a investičního optimismu či pesimismu. A právě tento psychologický aspekt má na tvorbu Elliottových vln největší vliv. Podobně jako Dowova teorie popisují akciový trh jako celek [39].

Formace označovaná jako Elliottova vlna má specifický průběh, první motivační fáze je složena z pěti menších růstových vln (1–5) (obrázek 4). Tato fáze je poté vystřídána korekční fází, která se skládá ze tří menších vln (A-C). Toto rozdělení by navíc mělo být dodrženo také v každém jednotlivých menších trendů. To je teoretický, idealizovaný tvar, rozlišit v praxi jednotlivé menší vlny bývá často komplikované. Měly by však při interpretaci být dodrženy tři základní pravidla [29]:

- Vlna 2 nesmí prolomit minimum vlny 1.
- Vlna 3 nesmí být nejkratším ze tří po sobě jdoucích stoupajících trendů.
- Vlna 4 nesmí zasáhnout zpět do vlny 1.



Obrázek 4: Elliottovy vlny

Zdroj: [29]

Techničtí analytici odhadují vývoj také pomocí studia mnoha dalších grafických formací, které se v průběhu času opakují na grafech akciových kurzů. Určité standardizované tvary pak vypovídají o budoucím vývoji. Využívá se také hranic podpory (Support Level) a hranic odporu (Resistance Level) nebo trendových kanálů a trendových linií. Prolomení stávající trendové linie informuje o změně trendu v opačný. Metody jsou popsány v [39] a [33]. Tato porovnání doplňují ostatní metody, pro tuto práci ale není jejich popis důležitý.

1.3.3. Indikátory technické analýzy

Veškeré metody a nástroje technické analýzy jsou zpravidla rozdělovány do dvou základních skupin. Jsou to grafické metody neboli charting a technické indikátory [39]. **Klouzavé průměry a metody na nich založené** jsou první skupinou technických indikátorů. Patří k nejoblíbenějším a nejpřesnějším metodám technické analýzy. Jedná se o typickou „following“ metodu neboli metodu následující trend [39]. Metoda je sice schopna informovat o změně trendu na trhu, nicméně vždy s určitým zpožděním. Metody jsou založeny na kalkulaci průměrného kurzu za několik období po sobě. Soubor průměrovaných kurzů se postupně obměňuje tím způsobem, že při uzavření nového období a vypočtení nejnovějšího průměru vypadne ze souboru ten nejstarší. Výsledná křivka se zobrazuje přímo v grafu kurzu a ze vzájemných interakcí vyplývají nákupní signály. K sestrojení křivky se využívá například jednoduchého, váženého, exponenciálního průměru a dalších metod.

Investoři si často výpočet použitý klouzavé průměry uzpůsobují podle svých potřeb. Především se mění období, za které je průměr vypočítáván, a tím i citlivost ukazatele. Často používanými intervaly jsou pětidenní, desetidenní, či dvacetidenní a padesátidenní průměry. Výjimkou ale není použití ani intervalu sta či dvou set dní. Oblíbené je také odvození od Fibonacciho posloupnosti, tedy průměrování v intervalech po 2, 3, 5, 8, 13, 21 nebo 34 dnech. Práce se ve druhé části zabývá porovnáním těchto metod s výstupem analýzy sentimentu, zvoleny byly takové intervaly, které jsou hojně používané v praxi a zároveň aplikovatelné v krátkém období. Nejjednodušší variantou je **SMA – Simple Moving Average**, tedy jednoduchý klouzavý průměr. Každý kurz získaný během zvolené periody má stejnou váhu, což znamená, že se nerozlišují kurzy starší a kurzy novější [33]. V této práci je použit třídenní, pětidenní a desetidenní klouzavý průměr, tedy $n = 3$; $n = 5$; $n = 10$. V případě výpočtu SMA se používá následující jednoduchý aritmetický průměr:

$$SMA = \frac{(P_1 + P_2 + \dots + P_n)}{n}, \quad (1)$$

kde $P_{1,2,\dots,n}$ = zavírací cena n -intervalu obchodních dnů,

n = počet dnů použitých pro klouzavý průměr.

Druhým příkladem je **EMA – Exponential Moving Average**. Exponenciální klouzavý průměr je obecně oblíbenější, a to hlavně kvůli tomu, že linka klouzavého průměru přisuzuje větší váhu nejnovějším cenám a dokáže tak rychleji zaznamenat změny trendu [32]. V této práci je použit pětidenní exponenciální klouzavý průměr, tedy interval $n = 5$. EMA lze vypočítat dle následujícího vzorce [9]:

$$\text{EMA} = \frac{p_1 + (1-\alpha)p_2 + (1-\alpha)^2 p_3 + \dots + (1-\alpha)^{n-1} p_n}{1 + (1-\alpha) + (1-\alpha)^2 + \dots + (1-\alpha)^{n-1}}, \quad (2)$$

kde $\alpha = \frac{2}{n+1}$,

$P_{1,2,\dots,n}$ = závírací cena n -intervalu obchodních dnů,

n = počet dnů použitých pro klouzavý průměr.

Další metodou je **MACD** neboli **Moving Average Convergence Divergence**. Je považován za jeden z nejspolehlivějších indikátorů [33]. Konstrukce toho indikátoru využívá několika klouzavých průměrů, zároveň ale vykazuje typické vlastnosti oscilátorů. Počítá se odčítáním dlouhodobého (typicky dvacetišestidenního) klouzavého průměru od krátkodobého (typicky dvanáctidenního). Do grafu je zakreslena i tzv. spouštěcí úroveň (typicky devítidenní SMA), která generuje nákupní/prodejní signály. MACD osciluje kolem nulové linie a pokouší se zachytit odchýlení od trendu a rostoucí pravděpodobnost změny trendu [39]. Intervaly typické pro MACD však vyžadují data za delší období, než má tato práce k dispozici. Jedná se ale o natolik rozšířený indikátor, že ho nelze opomenout, byť nebude přímo použit.

Prvním velmi často užívaným ukazatelem z řady **oscilátorů** je **RSI – Relative Strength Index**. Index relativní síly měří vnitřní sílu na trhu a jeho hodnoty se pohybují v rozmezí 0 až 100 [11]. Byl zkonstruován za účelem odstranění obecně známých nedostatků oscilátorů a slouží k identifikaci trendů a k indikaci signálů ke vstupu na trh či jeho opuštění [33]. Výpočet RSI je následující [11]:

$$\text{RSI} = 100 - \frac{100}{1 + \text{RS}}, \quad (3)$$

kde RS = součet kladných cenových změn za dané časové období / součet záporných cenových změn za dané časové období.

Nejčastěji používaným časovým rozpětím je 14, a dále pak 9 a 26. Pro potřeby této práce bylo zvoleno rozmezí 9 dnů. Čím kratší je časové období, tím více indikátor osciluje a dává více signálů. Jeho hlavním přínosem je zjištění, kdy je trh tzv. překoupený nebo předprodaný. Důležité hodnoty jsou hladiny 30 a 70. Pokud se RSI pohybuje na hodnotách 30 a níž, naznačuje nám, že je trh předprodaný a může dojít k obratu směrem nahoru. Naopak, pokud se hodnota RSI pohybuje na hodnotě 70 a více je trh překoupený [11]. Je možné zvolit pásmo nejen 30 / 70, ale i 20 / 80 (pro volatilní tituly) a 40 / 60 (pro indexy) [6].

Druhým zástupcem je **CCI – Commodity Channel Index**. Ten měří pohyby ceny kolem jejího statistického průměru. Pokud se CCI pohybuje v úrovni vysokých hodnot (> 100), znamená to, že ceny jsou neobvykle vysoko nad svojí průměrnou hodnotou. Pokud se CCI pohybuje na nízkých hodnotách (< -100), potom to znamená, že ceny jsou příliš nízko pod svojí průměrnou hodnotou (překročení hodnot ± 100 může značit silný trend a hodnoty ± 200 naznačují překoupenost a přeprodanost). Vzorec pro výpočet CCI [11]:

$$CCI = (P - SMA) / (0,015 \times D), \quad (4)$$

kde P = zavírací cena,

SMA = zvolený klouzavý průměr,

D = normální odchylka.

Další hojně využívanou možností je **Stochastický oscilátor – SO**. Udává, kde se nachází poslední uzavírací kurz vzhledem k cenovému rozpětí. 0 % odpovídá pětidennímu minimu a 100 % maximu. Signálem je protnutí nastavené procentní hranice 20 / 80 nebo 25 / 75 nebo 30 / 70 [6]. Vzorec Stochastického oscilátoru vypadá následovně [6]:

$$\%K = (C - L14 / H14 - L14) * 100, \quad (5)$$

kde C = nejnovější zavírací cena,

$L14$ = nejnižší obchodovaná cena za posledních 14 obchodních dní,

$H14$ = nejvyšší obchodovaná cena za posledních 14 obchodních dní,

$\%K$ = současná hodnota Stochastického oscilátoru [12].

1.4. Psychologická analýza

Psychologická analýza má pro účely této práce zvýšenou důležitost. Pojem sentiment spadá právě do této oblasti, má blízko k pojmu emoce a subjektivita v téhle oblasti je tím hlavním důvodem existence finančních trhů. Pokud by si každý člověk myslel ve stejný okamžik to samé, co ostatní, trhy, jak je známe v dnešní podobě, by vůbec neexistovaly, protože by nikdo neměl zájem o směnu.

Předmětem zkoumání psychologické analýzy tedy není cenný papír samotný, nýbrž člověk investor, tedy lidský faktor v procesu investování [39]. Vychází z toho, že investiční rozhodnutí je ovlivněno emocemi. Touha po zisku, lakota a hrabivost lidí zatlačují racionální úvahy do pozadí. Na trhu pak často převládají faktory neracionálního charakteru [16].

1.4.1. Příslušnost k davu

„Osobitý“ přístup k investičnímu rozhodování občas svádí investory vydat se jedním směrem, zatímco jejich „stádový instinkt“ je navádí směrem přesně opačným. Toto rozpolcení osobnosti při rozhodování je ve skutečnosti u všech jedinců přirozeným důsledkem vnitřního vztahu mezi integritou jedince a členstvím ve skupině [31]. Hlavním požadavkem pro členství ve skupině je přitom povinnost přijmout určité vzorce chování, které jsou v souladu se základními cíli dané skupiny. Je třeba si přitom uvědomit, že není ničím neobvyklým, když se lidé, poté co začnou patřit do nějaké nové skupiny, začnou chovat naprosto rozdílně [31]. Dochází k potlačení povahových vlastností jednotlivců a k setření jejich intelektuálních schopností [16].

Jedním z prvních lidí, kteří se hlouběji zabývali fenoménem lidských skupin, byl Gustav Le Bon. Východiskem jeho teorie je analýza a rozbor celkové povahy davu, kterou nazývá kolektivní duší [39]. Jeho zásadní dílo s názvem Psychologie davu vzniklo v roce 1895 a dodnes je považováno za jedno z nejvýznamnějších děl v oboru sociální psychologie. Jeho teorie byly následně opakovaně potvrzovány dalšími psychoanalytiky, mezi něž patřili například Sigmund Freud nebo Carl Jung, ale také teoretiky, jako byl Arthur Koestler. Le Bon považoval dav spíše za psychologický nežli fyzický fenomén a zastával názor, že jakýkoli počet jinak zcela nezávislých a prostorově oddělených jedinců může vytvořit dav za předpokladu, že všichni jeho členové mají nějaký společný důvod [31]. Jako příklad takových skupin lze uvést například firmy, fotbalová družstva, armády, náboženské sekty a příslušníky národů, ale stejně tak třeba pochody demonstrantů. Schopnost nepodlehnout davové psychóze mají pouze silné osobnosti [16].

Stěžejní je tedy fakt, že účastí v davu dochází u každého jednotlivce ke změně chování. Jedincem je i individuální investor na finančních trzích, je základním stavebním kamenem hierarchie na finančních trzích, na situaci na trhu se aktivně podílí a je tedy součástí davu. Pokud ale investor podlehne davu, nemůže být lepší ani horší než průměr [16]. Každý investor má specifickou investiční strategii, přesto existují jen dvě možnosti, na kterých může každý z nich v danou chvíli vydělat. Jednou z nich je očekávaný růst cenového trendu, druhou možností je potom pokles cenového trendu. Drasnarova koncepce vysvětluje pohyby cen akcií jako důsledek působení dvou protichůdných vlastností. Jde o chamtivost a strach [33]. Dav investorů, který očekává růst, se nazývá býčím, dav, který očekává pokles, se nazývá medvědí [31]. Stejně tak lze rozdělit veškeré investory do davu, který je v danou chvíli v úspěšných, tedy ziskových pozicích a na dav neúspěšných, tedy těch ve ztrátových pozicích. A přesto, že se investoři fyzicky nekoncentrují na jednom místě, tvoří dav, který podléhá prvkům davové

psychologie jako každý jiný [31]. Dnes tím více kvůli rozšířenosti telekomunikace se skrz televizi, noviny, telefony, internetové zpravodajství a sociální sítě tyto investoři navzájem ovlivňují a tyto kanály dnes plně nahrazují fyzický kontakt běžný v jiných davech. Psychologickou reakci od investora vyvolá každá nová informace a ty jsou dnes dostupné ve větším množství než kdy dříve.

1.4.2. Vliv emocí

Rozhodnutí nakoupit či prodat cenné papíry v nás pokaždé vzbuzuje určité emoce. Rozhodnutí o provedení obchodu sice může být uskutečněno na racionálním základě, ale samotný akt obchodování přináší i nutnost vložit do hry určitou finanční částku. Od té chvíle je v nás vzbuzen pocit, že nesmíme „udělat chybu“. Vývoj kurzu pak nevyhnutelně přinese pocity spokojenosti či nespokojenosti, ty budou mít za následek rovněž fyziologické změny v těle investorů – změní se srdeční tep a rychlost dýchání, začnou se potit dlaně a podobně. Výše popsané pocity se navíc dále zintenzivňují, pokud se jednotlivec spojí s dalšími lidmi [31].

Bing Liu definuje emoce spolu s dalšími pojmy [23]:

- Afekt – Pocit nebo emoce, obzvláště vyjádřena výrazem tváře či řečí těla.
- Emoce – Mentální stav, který nastává spontánně, bez záměrné snahy. Obvykle je doprovázen fyziologickými změnami.
- Nálada – Stav mysli či emoce.
- Pocit – Afektivní stav vědomí, zapříčiněný například emocemi, sentimentem či touhami.

Pokud se obchodní pozice vyvíjí správným směrem, kromě dobrého pocitu z vydělaných peněz a zvýšeného sebevědomí najednou investor ucítí potřebu podělit se o svůj úspěch s ostatními investory, kteří drží podobné obchodní pozice [31]. Tyto konverzace s ostatními lidmi utvrdí nejen jejich přesvědčení, že je jejich obchodní pozice správná, ale také že je správný i samotný proces rozhodování, který jejich investici předcházel. Investoři mohou pocítit další uspokojení, pokud se jim dostanou do rukou například různé novinové články či burzovní zprávy, které jim přinesou další důkazy potvrzující správnost jejich rozhodnutí. A co víc, členové „úspěšného“ davu budou mít tendenci zdůrazňovat slabost argumentů „neúspěšného“ davu [31]. Emoce je spojena s určitým objektem, například osobou, událostí, věcí nebo tématem. Bývá intenzivní a zaměřená velmi specificky. Typicky také má také krátké trvání [23]. Bez ohledu na to, jak racionální bylo původní rozhodnutí o vstupu do určité obchodní pozice, samotný akt zahájení

obchodování uvrhne jedince do prostředí, kde se lidé přestávají řídit výhradně rozumem a kde je velmi snadné podlehnout nátlaku davu [31].

A kde dnes snáze nahlédnout na tisíce názorů ostatních investorů než na internetu. Členové ať už úspěšného nebo neúspěšného davu zde mohou svůj názor navíc vyjádřit anonymně, pod libovolnou zvolenou přezdívku. Takové zaštitění internetovou clonou často způsobí, že investoři vystupují agresivněji, než jak by tomu bylo při osobním kontaktu. Razantněji kritizují jiné názory a bezohledněji obhajují ty své, jelikož nejtvrďší reakcí na jejich kritiku mohou být maximálně jiné agresivní příspěvky, které není nijak složité ignorovat.

1.4.3. Názor

Před definováním sentimentu jako cíle naší analýzy je vhodné definovat také názor z jednoho prostého důvodu. Názor je předmětem výzkumu větve Data miningu s názvem Opinion mining a pojem názor zahrnuje jak námi hledaný sentiment, tak hodnocení, ocenění nebo přístup a dále přidružené informace jako názorový cíl a identifikaci osoby, která tento názor vyjadřuje. Sentiment je tedy pouze jednou ze složek názoru a vyjadřuje pouze pozitivní či negativní pocit, který tento názor implikuje. Jedna z definic názoru je tedy vyjádřena následující čtveřicí:

$$(g, s, h, t), \tag{6}$$

kde g je názorový cíl, s je sentiment názoru ve vztahu k cíli g , h je nositel názoru (osoba nebo organizace, které názor vyjadřuje) a t je čas, kdy byl názor vyjádřen [23].

Tyto čtyři komponenty jsou zcela zásadní pro vyjádření názoru. Pokud jeden z nich chybí, nastane problém se správným pochopením názoru. Například komponent času je v praktickém životě velmi důležitý, neboť udává kontext názoru. Názor formulovaný před dvěma lety není stejný, jako tento názor formulovaný dnes [23].

1.4.4. Sentiment

Sentiment obecně je definován jako základní pocit, přístup, ohodnocení nebo emoce spojená s názorem. Je definován následující trojicí:

$$(y, o, i), \tag{7}$$

kde y je označení pro typ sentimentu, o je orientace sentimentu, a i je intenzita sentimentu [23]. Typ sentimentu může být určen několika způsoby. Existují klasifikace založené na lingvistickém přístupu, psychologickém přístupu a přístupu založeném na zákaznických průzkumech. Orientace sentimentu může být pozitivní, negativní nebo neutrální. Neutrální obvykle znamená absenci sentimentu. Intenzita sentimentu určuje sílu vyjádření. Každý z typů

sentimentu může být ještě vyjádřen s různým stupněm intenzity [14]. Lidé to mohou vyjádřit dvěma způsoby. Prvním je užití přímo výrazu s přiměřenou silou, například slovo „vynikající“ nese silnější sentiment než „dobrý“. Druhým způsobem je užití dalších slov, která přidají na intenzitě původního výrazu, například „velmi dobrý“ [17].

V praxi se často pro měření používá nějaká diskrétní hodnotící škála, například hodnocení pěti úrovněmi – hvězdičkami. To se poté z hlediska sentimentu dá interpretovat následujícím způsobem [23]:

- *Emočně pozitivní (+2 nebo 5 hvězd);*
- *racionálně pozitivní (+1 nebo 4 hvězdy);*
- *neutrální (0 nebo 3 hvězdy);*
- *racionálně negativní (-1 nebo 2 hvězdy);*
- *emočně negativní (-2 nebo 1 hvězda).*

Tato data ale dostaneme pouze, když se uživatel rozhodne zareagovat a hvězdičky vyplnit. Navíc je to velmi omezený pohled, který neumožňuje přesně identifikovat jednotlivé hodnocené entity nebo atributy.

V tržním prostředí vyvolá každá informace investorovu reakci. Nemusí se jí řídit, ale nějaké emoce díky ní vyplují na povrch, ať už si to přeje či nikoli. Emoce vyústí v nějaké celkové cítění situace na trhu, toto cítění označujeme, hovoříme-li o sentimentu na finančních trzích. Pozitivní či negativní očekávání. Investor si často tento sentiment i plně uvědomuje, snaží se jeho správnost poté potvrdit nebo vyvrátit jinými metodami. Následuje několik definic sentimentu přímo od investorských webových portálů.

Podle portálu www.fxstreet.cz [7]:

„Market sentiment (tržní sentiment) - nálada na trhu, psychologie davu, měřítko bullish (býčích) nebo bearish (medvědí) postojů mezi investory a obchodníky.“

Definice podle portálu www.xtb.com [40]:

„Tržní sentiment může být definován jako celkový postoj investorů k určitému aktivu nebo k trhu jako celku. Je to nálada na trhu, která se promítne do ceny aktiva. Je důležité si uvědomit, že sentiment není vždy založen na fundamentech.“

Měřit nějakým způsobem sentiment je velmi složité, existují indikátory, které pro tuto kvantifikaci vyčíslují počty a poměry otevřených bých a medvědích pozic, tato práce se ale zabývá extrakcí sentimentu přímo z psaného textu, proto zde nejsou popsány.

1.5. Analýza textu

Sedm hlavních přístupů k získávání informací z textu bylo vyčleněno v publikaci [26]. Přístupy se v praktické aplikaci často překrývají a doplňují. Jedná se o:

- *Search and Information Retrieval (IR)* – vyhledávání a získávání informací,
- *Document Clustering* – klastrování (shlukování) dokumentů,
- *Document Classification* – klasifikace dokumentů,
- *Web Mining* – dolování dat z webu,
- *Information Extraction (IE)* – extrakce informací,
- *Natural Language Processing (NLP)* – zpracování přirozeného jazyka,
- *Concept Extraction* – extrakce konceptu.

Vyhledávání a získávání informací (IR) je využití klíčových slov a indexování k prohledání rozsáhlých databází. Na tomto principu fungují například internetové vyhledávače. Klastrování dokumentů je přístup, který, za použití algoritmů data miningu, vysleduje podobnost určitých dokumentů. Na základě této podobnosti potom sestaví požadované shluky. Klasifikace dokumentů využívá systémového učení. Na rozřazené databázi dokumentů natrénuje postup, kterým pak přiřazuje nezatříděné dokumenty do jednotlivých tříd. Dolování dat z webu se týká textu na internetu. Ten se svou strukturou a provázaností s jinými internetovými texty výrazně liší od běžných textů, proto je vyčleněn jako samostatný přístup. Extrakce informací (IE) je náročnější technika v oblasti analýzy textu, která vyžaduje použití pokročilých algoritmů a software. Cílem této metody je vytáhnout strukturovaná data z nestrukturovaného textu. Zpracování přirozeného jazyka (NLP) je postup, kterým se stanoví vstupní proměnné pro ostatní data miningové metody. Dříve byla tato metoda vnímána spíše jako lingvistický nástroj, časem se ale začala uplatňovat také v informatice a dnes už je neodlučně spojena i s analýzou textu [8]. Extrakce konceptu analyzuje jednotlivá slova a fráze dokumentu a třídí je do sémanticky – neboli významově – podobných skupin. Snahy automatizovat tento přístup však zatím nemají příliš přesné výsledky, neboť vyžadují vysokou míru porozumění kontextu.

Tato práce se zabývá získáváním informací z webu. Bližší kategorizaci nabízí ve svém článku Feldman [5]. Analýza sentimentu nebo Opinion mining, doslovně přeloženo jako dolování názorů, má za úkol zjistit názor autora textu na určité téma.

Proces rozhodování a utváření názoru je u každého jedince ovlivňován jednak názory vůdčích osobností a dále také názory ostatních, „běžných“ lidí. Při nákupu zboží online například typicky zákazník začne průzkumem recenzí aneb názorů ostatních uživatelů. Stejným způsobem postupuje investor při rozhodnutí o nákupu nebo prodeji akcií, snaží se maximálně zorientovat v dostupných informacích, názory ostatních investorů nevyjímaje, a poté u něj převládne pozitivní nebo negativní sentiment. V posledních letech se na tuto oblast zaměřuje pozornost jak drobnějších startupových společností, tak velkých vývojářů jako SAS, SPSS. Možnost rychle analyzovat velké množství textu a zjistit, jaké názory obsahuje, může být nedocenitelnou výhodou ve všech oborech. Sociální sítě jsou zlatý důl pro společnosti, které chtějí sledovat a hodnotit svou reputaci a znát zpětnou vazbu na konkrétní nově zavedené výrobky nebo aktivity. Marketing, public relations, řízení projektů, politika, investování a obchod, tam najde technologie analýzy sentimentu největší uplatnění. Může přitom být odlišný cíl použití i zdroj dat, lze zde tedy vyčlenit několik úrovní užití analýzy sentimentu [5]:

- Analýza sentimentu na úrovni dokumentu – nejjednodušší užití. Předpokládá, že autor dokumentu hodnotí jednu konkrétní entitu a podle souboru pravidel stanoví výsledný celkový sentiment. Pravidla jsou stanovena na základě práce s testovacím souborem, využitím balíčků POS (Part of Speech) aneb lexikonu [8]. Výsledek je vyjádřen pro celý dokument na základě počtu pozitivních, negativních, případně neutrálních výrazů buďto numerickou hodnotou nebo například hodnotící škálou [5]. V praxi může výstup analýzy vypadat tak, že v článku pojednávajícím o novém výrobku firmy F je obsažen převážně kladný sentiment, hodnocení lze převést na 4 z 5 hvězdiček.
- Analýza sentimentu na úrovni vět – pracuje s faktem, že v textu může být zmíněno více entit aneb obsaženo několik odlišných názorů na jednu entitu [8]. Pro aplikaci je však nutné předem znát entity v textu obsažené. Druhým úskalím je předpoklad, že jedna věta vyjadřuje jeden názor, což nemusí být vždy pravda. Některé metody také usilují o extrakci objektivních faktů a následně analýzu pouze subjektivně zabarvených vět. Tak jako tak je to už bližší a propracovanější pohled než analýza celého dokumentu, výsledkem jsou shluky dle témat jednotlivě ohodnocené na spektru negativní až

pozitivní [5]. Konkrétně si tímto způsobem může firma například zjistit, že s jejími službami A a B jsou klienti spokojeni, ale služba C je vnímána negativně.

- Analýza sentimentu jednotlivých aspektů – nejpodrobnější pohled, kde může být v rámci každé jedné věty sledováno hned několik věcí. Uživatel se například vyjádří k několika vlastnostem (aspektům) produktu (entity) od modelové firmy. Kladně bude hodnotit kompaktnost a cenu, negativně materiál a výdrž baterie. Výsledný sentiment na úrovni věty by se mohl tvářit neutrálně, neboť kladné i záporné dojmy se „vyváží“, přesto sdělení obsahuje pro firmu klíčové informace o tom, co uživatelé subjektivně vnímají jako silné a slabé stránky. V analýze se pro identifikování aspektů používá extrakce podstatných jmen. Ta, která se v datovém souboru vyskytují častěji, než je stanovená hladina, se stávají jednotlivými aspekty [5]. Často ale není v textu aspekt přímo vyjádřen podstatným slovem, místo toho je popsán přídavným jménem či určitým ustáleným výrazem. Například v hodnocení: „Tento telefon je příliš těžký“ je negativně hodnocen aspekt váhy, přestože slovo váha zde nefiguruje. Aspekt tedy není vyjádřen explicitně, ale implicitně. Tehdy je třeba přiřadit toto implicitní vyjádření k explicitnímu aspektu, autor Bing Liu pro řešení tohoto problému uvádí dvoufázové asociační pravidlo společného výskytu. Tato práce se ale detailním průzkumem jednotlivých aspektů nezabývá.
- Komparativní analýza sentimentu – v analyzovaném textu je hodnocen shodný aspekt u více entit, uživatel například vyjadřuje sentiment sdělením, že výrobek firmy A je rychlejší, odolnější nebo více náročný na údržbu, než podobný výrobek od společnosti B. Zde je cílem analýzy identifikovat věty, které obsahují srovnání a vybrat tu entitu, která je autorem hodnocení preferovaná. Pro tento nesnadný úkol sestavili autoři Jindal a Liu [17] relativně malý list slov, která ale pokryjí 98 % názorů obsahujících srovnání. Díky tomuto nástroji má analýza vysokou odezvu, ale stále poměrně nízkou přesnost. Pro vyfiltrování vět bez srovnání se dále používá data miningový algoritmus CSR (class sequential rule), který identifikuje jednotlivé entity na základě použitého typu srovnání a také přítomnosti negací.
- Analýza sentimentu v závislosti na slovníku – není tak úplně samostatnou úrovní analýzy, bývá součástí analýz různých úrovní popsaných výše, použitý slovník je totiž zcela zásadním prvkem pro většinu algoritmů. Slovník je sestaven buďto manuálně, to

ale obsahuje vysokou míru subjektivity, navíc sestavit vhodný slovník pro každou analýzu zvlášť by bylo velmi náročné. Další možností je použití obecného jazykového slovníku, tento způsob zase nemusí pojmout slova specifická v analyzovaném odvětví nebo hovorové výrazy, v nichž mají určitá slova jiný význam než v běžném jazyce. Také se využívá slovník obsahující pouze kořeny slov. Ten lze dále upravit kombinací s lexikální databází synonym a antonym například WordNet. Algoritmus využívající zmíněnou síť slov je uveden například v [19]. Metoda definuje vzdálenost $d(t_1, t_2)$ mezi pojmy t_1 and t_2 jako délku nejkratší cesty mezi t_1 and t_2 v síti WordNet. Orientace t je definována jako:

$$SO(t) = \frac{d(t, negativní) - d(t, pozitivní)}{d(pozitivní, negativní)}, \quad (8)$$

kde $SO(t)$ je síla sentimentu t , $SO(t) > 0$ značí pozitivní t , jinak je t negativní. Jiné algoritmy využívají například zásady konzistence sentimentu a neznámým adjektivům, která stojí ve větě vedle známých adjektiv a používají určité spojovací výrazy, přiřazují stejný sentiment. Na začátku je tedy slovník, ve kterém je přídavné jméno „výkonný“ spojeno s pozitivním sentimentem. Pokud uživatel v hodnocení napíše „tento telefon je výkonný a lehký“ lze díky skladbě věty a spojce „a“ přiřadit pozitivní sentiment i slovu „lehký“ a klastr pozitivních slov ve slovníku o něj rozšířit [5].

Zatím se v oblasti analýzy sentimentu hovoří o jednotlivých metodách a algoritmech, jsou zde ale pochopitelně silící tendence zdokonalovat a komplexní řešení se moci už nazývat spíše umělou inteligencí. Dále jsou uvedeny nejčastější problémy, které analýzu sentimentu dnes komplikují.

1. V první řadě posuzování jednotlivých slov nevystihne sentiment celé věty dost přesně, pro to je třeba vyvinout lingvisticky dokonalejší postupy [5].
2. Dalším problémem je to, že každá entita nebo aspekt může být nazývána různými jmény. Výrazy jako „životnost baterie“ nebo „využití energie“ mohou mít v rámci hodnocení stejný smysl, ale je v nich vyjádřen jinými slovy. Algoritmus, který dokáže u takových výrazů rozpoznat, že hovoří o stejném aspektu stejné entity, by musel být velice citlivý. Toto je pochopitelně problematické i druhým směrem, tedy užitím synonym. Pokud jedno pojmenování může označovat různé entity, není možné je jednoduše automaticky rozklíčovat [23].

3. Pokud dokument pojednává o několika entitách současně, neexistuje v současnosti dostatečně přesný způsob, jak správně identifikovat relevantní text ke každé entitě [5].
4. Velmi problematické je užití sarkasmu, zatím není v analýzách ošetřeno prakticky vůbec [22]. Například výraz „to je opravdu super“ bude z hlediska sentimentu ohodnocen pozitivně, přestože může v určitém kontextu vyjadřovat negativní pocity.
5. Náročné na analyzování jsou také texty s velkým obsahem šumu, tedy texty gramaticky nesprávné, se slangovými výrazy a bez diakritiky. To jsou velmi typické rysy příspěvků v internetových diskuzích. Mezi uživateli takových diskuzí je také oblíbené vyjadřovat svůj postoj vložením odkazu na jiný článek či obrázku, oba případy nesou určitý sentiment, který ale analýze textu unikne [22].

1.5.1. Slovníky

Zde zmíněný sentiment bývá označován také jako tón textu. Ve většině postupů k jeho určení hraje klíčovou roli použitý slovník, různí autoři pracují s různými verzemi.

Bing Liu

Kromě metody vzdáleností od [19] popisuje Bing Liu ještě několik dalších metod stavby slovníku. Nejčastěji je nejprve manuálně sestaven malý soubor slov nebo kořenů slov, u kterých je sentiment známý podle obecných slovníků. Poté je tento soubor rozšiřován v postupných krocích algoritmem, který hledá synonyma či antonyma slov z původního souboru v dalších slovnících. Po připsání nově nalezených slov do souboru začíná nová iterace. Iterace končí, není-li nalezeno žádné nové slovo [14]. Po dokončení procesu je provedeno čištění souboru od chyb. To lze provést manuální prohlídkou celého souboru. Čištění lze také provést přiřazením hodnoty vyjadřující sílu sentimentu ke každému slovu a následnou aplikací pravděpodobnostních metod [22].

Tim Loughran, Bill McDonald

Různé verze však často mají, tak jako tato, společný základ v Harvardském psychologickém slovníku s označením Harvard-IV-4 TagNeg (H4N). Jednou z hlavních výhod jeho použití je objektivita, výzkumník totiž nemůže skladbu slovníku subjektivně ovlivnit. Pro některé texty tento slovník funguje dobře, otázkou ale je, jak moc je vhodný pro použití ve specifických oblastech. Autoři [24] přinesli důkaz o špatné klasifikaci řady slov při použití slovníku v analýze v oblasti financí. Podle nich 73,8 % slov negativně označených v Harvardském slovníku není typicky negativních ve finančním kontextu. Slova jako *tax*, *cost*, *capital*, *board*,

liability, foreign, (daň, náklad, kapitál, závazek, cizí) jsou obsaženy v Harvardském seznamu negativ [24]. Tato slova se objevují s vysokou frekvencí ve finančních souhrnných výročních zprávách o výkonu společnosti, označovaných 10-Ks, obvykle ale neznamenají více než body z jednání představenstva a viceprezidentů společnosti. Dále pak slova jako *mine, cancer, crude (oil), tire, or capital, (důl, rakovina, surová (ropa), pneumatika* či *kapitál)* pravděpodobněji identifikují specifický průmyslový segment než negativní finanční události. Autoři [24] tedy pozměnili seznam negativních slov Harvardského slovníku. U slov, která se vyskytla v alespoň 5 % zpráv 10-Ks, zkoumali přesný význam ve světě financí. Ohlášené finanční výsledky společností pak měly výraznou souvislost s výstupem analýzy za použití tohoto seznamu. Autoři dále doporučují zapracovat do analýzy systém vah, kterým se sníží nežádoucí šum a zvýší dopad méně frekventovaných slov na výsledek.

Elaine Henry

Autorka hodně pracuje s myšlenkou, zdali je sentiment do textu umísťován záměrně s cílem propagovat či manipulovat, zaměřuje se ale na články z tisku a výroční zprávy. Tón neboli sentiment definuje jako působení komunikace. Účelový propagační článek bude mít pochopitelně pozitivní tón, protože autor jím chce čtenáře ovlivnit. Obecně ale pozitivní tón nemusí být vždy účelový. Autorka mapuje strukturu takových článků, cílí na frekventovaná slova většinou s komparativní funkcí a svůj slovník tomu přizpůsobuje.

1.5.2. Specifika debat a komentářů

Obecně na sociálních médiích lze nalézt dva typy textového obsahu: samostatně stojící příspěvky, například uživatelské recenze a blogy, a online dialogy, debaty a diskuze. Online dialogy jsou konverzace, která typicky zahrnuje interaktivní výměny mezi dvěma nebo více zúčastněnými. Na rozdíl od samostatných příspěvků, které jsou typicky nezávislé jeden na druhém. Online dialogy jsou obvykle plné názorů. Kromě negativního a pozitivního sentimentu také obsahují souhlasy a nesouhlasy, které lze chápat jako interaktivní formu sentimentu [5]. S přihlédnutím na tyto uživatelské interakce mohou být provedeny další analýzy. Lze například objevit postoj jednotlivých diskutujících a na základě něj je přiřadit k určitému ideologickému davu. Těžením výrazů obsahujících souhlas či nesouhlas lze také odhalit problémový nebo kontroverzní obsah a zároveň argumentační povahu diskutujících [27].

2. Analýza sentimentu sociálních médií

2.1. Příprava dat

Vzhledem k neuvěřitelnému kvantu informací, které díky dnešnímu využívání internetu a rozšíření sociálních médií produkujeme, je v první řadě nutné vyhledat a vybrat tu správnou síť pro analýzu.

2.1.1. Sociální média

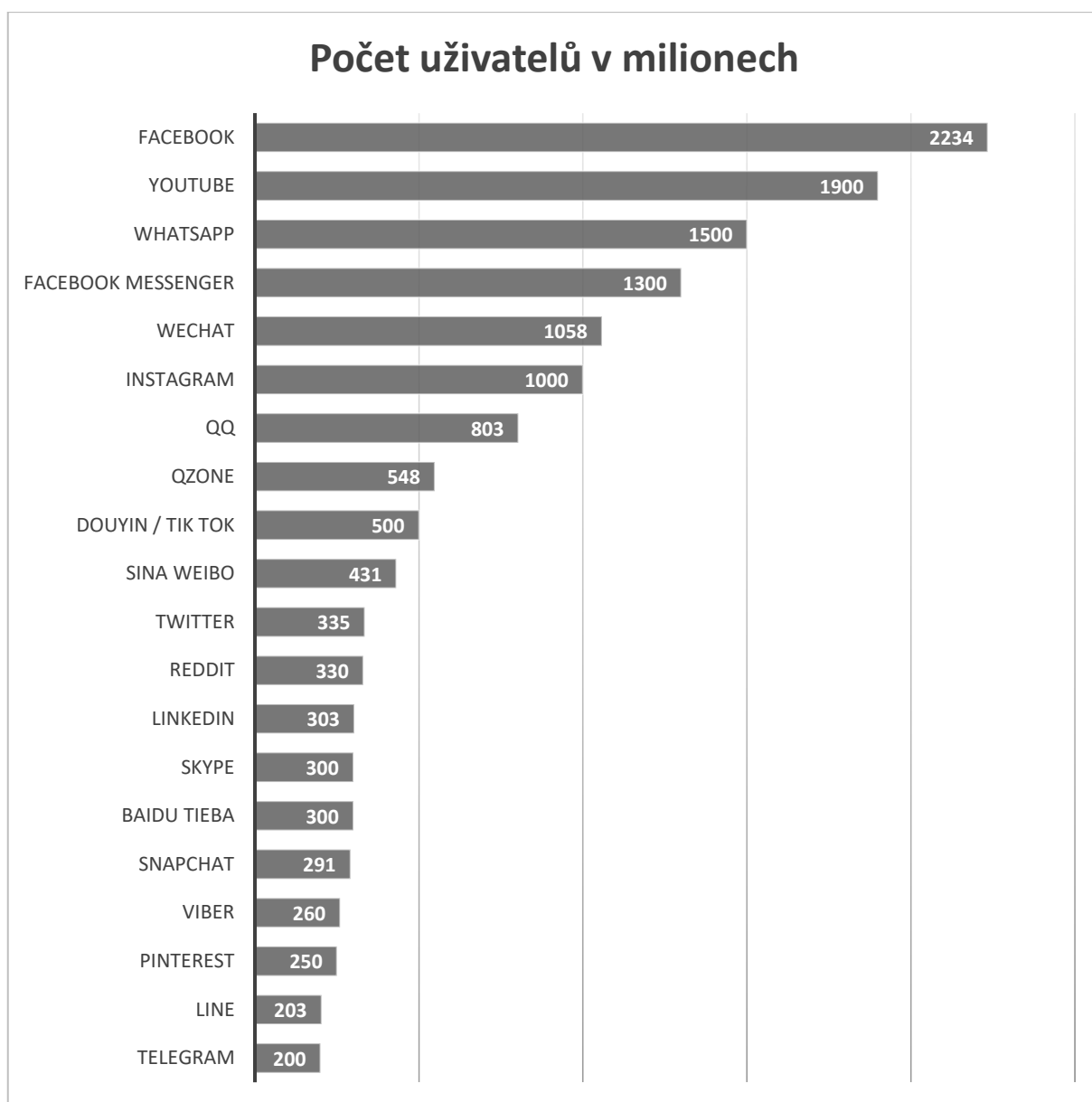
V březnu roku 2018 zpracoval portál Forbes statistiky užívání internetu a počet dat vytvořených lidstvem za den spočítal na 2,5 trilionů bytů, tedy 2,5 exabytu. Zde je výčet několika faktů, uvedená čísla navíc v posledních dvou letech vzrostla o 7,6 % [25]:

- Více než polovina vyhledávání na internetu je zadána z mobilního telefonu.
- Internet využívá více než 3,7 miliard lidí.
- Google průměrně uskuteční více než 40 000 vyhledávání za sekundu, což je 3,5 miliardy vyhledávání denně.
- Google provádí 77 % všech vyhledávání, celkově se ve všech vyhledávacích každý den vyhledává přes 5 miliard dotazů.

Pokud se pozornost zaměří blíže na sociální média, zde jsou průměrná data, která uživatelé vytvoří za jednu minutu [25]:

- Uživatelé Snapchatu nasdílí 527 760 fotografií.
- Více než 120 profesionálů se přidá na síť LinkedIn.
- Uživatelé shlédnou 4 146 600 videí na YouTube.
- Na Twitteru je zveřejněno 456 000 „tweetů“.
- Uživatelé Instagramu nahrají 46 740 fotek.

Různé skupiny lidí mají v oblibě různá sociální média a jejich obsah se tedy může lišit. Bude-li chtít člověk pročíst komentáře k nejnovějšímu hudebnímu videu, může zamířit například na YouTube. To ale nebude pravděpodobně první volbou, bude-li chtít s přáteli z mého okolí sdílet nějaké zážitky, v tom případě zavítá spíše na Facebook. Pokud o sobě bude chtít dát vědět potenciálnímu zaměstnavateli a bude se chtít více dozvědět o společnostech v regionu, rozvine svoje aktivity naopak spíše na síti LinkedIn. Následující graf (obrázek 5) obsahuje nejpopulárnější síť na světě dle počtu aktivních uživatelů v říjnu roku 2018.



Obrázek 5: Sociální sítě dle počtu uživatelů

Zdroj: Vlastní zpracování podle [35]

Řada z těchto největších světových sociálních sítí není v českých luzích a hájích obecně příliš známá a používaná. Z toho vyplývá další věc, kterou bylo nutné při stanovení vzorku k analýze zvážit, je region. Ten musí korespondovat u investorů i příspěvateľů v sociálních médiích. Jako zdroj dat k analýze byla vybrána sociální síť Yahoo Finance, jelikož Yahoo je hojně vyžívaný portál v USA a také jeho „odnož“ zaměřená přímo na finance umožňuje odfiltrovat už v této počáteční fázi část šumu, který je na sociálních sítích a internetových diskuzích obecně velkým problémem. Jinak řečeno, u takto „specializované“ sítě existuje předpoklad, že zde bude více příspěvků k tématu od lidí, kteří se na finančních trzích aktivně pohybují. Z akciových titulů bylo v listopadu 2018 pro analýzu vybráno zastoupení dvaceti společností s největší tržní

kapitalizací, které splňují dvě podmínky. Zaprvé spadají pod region Spojené státy americké, zadruhé jejich volume, tedy počet obchodních kontraktů denně, je větší než 2 000 000. Obě tato kritéria slibují, že na Yahoo Finance bude k nalezení dostatečné množství dat k analýze. Data byly sbírána za období dvou měsíců, tedy od 1.11.2018 do 31.12.2018. Výsledkem je následující seznam:

Apple Inc. (AAPL); Microsoft Corporation (MSFT); Amazon.com, Inc. (AMZN); Facebook, Inc. (FB); Johnson & Johnson (JNJ); Alibaba Group Holding Limited (BABA); JPMorgan Chase & Co. (JPM); Exxon Mobil Corporation (XOM); Visa Inc. (V); Walmart Inc. (WMT); Bank of America Corporation (BAC); Pfizer Inc. (PFE); Wells Fargo & Company (WFC); Verizon Communications Inc. (VZ); The Procter & Gamble Company (PG); Chevron Corporation (CVX); AT&T Inc. (T); Intel Corporation (INTC); Cisco Systems, Inc. (CSCO); The Coca-Cola Company (KO).

Předpoklad, že aktivněji obchodované tituly mají také aktivnější diskuzi, se ale během fáze sběru dat ukázal jako chybný. V diskuzích u některých titulů se objevovaly několikedenní úseky bez příspěvků, tempo komentářů nemělo dostatečnou frekvenci, a proto nebylo možné takový datový soubor analyzovat. Ve finálním seznamu má tedy zastoupení celkem osm akciových titulů. Po časově náročné fázi čištění a úpravy dat je množství příspěvků, tedy jednotlivých analyzovaných případů, celkem 85 430. Zde je tedy finální seznam titulů seřazený dle četnosti příspěvků. Od nejpočetnějších po nejméně početné.

1	Apple Inc.	AAPL
2	Bank of America Corporation	BAC
3	Amazon.com, Inc.	AMZN
4	Cisco Systems, Inc.	CSCO
5	Facebook, Inc.	FB
6	Alibaba Group Holding Limited	BABA
7	AT&T Inc.	T
8	Walmart Inc.	WMT

2.1.2. Sběr a čištění dat

Na sociální síti Yahoo Finance jsou přehledné finanční i fundamentální informace k jednotlivým akciovým titulům, včetně veřejných online debat. Prvním krokem bylo

odstranění automatických filtrů z debat, díky kterým se nezobrazovaly všechny komentáře, ale pouze například nejvíce populární.

Diskuze na Yahoo Finance umožňuje přesně zobrazit den zveřejnění příspěvku pouze u příspěvku mladších než jeden měsíc. Některé diskuze jsou navíc tolik frekventované, že se v procesu sběru projeví i technická náročnost zpracování dat. Proto sběr dat probíhal v několika fázích se zhruba čtrnáctidenním rozmezím. Takto například vypadá jeden z komentářů:

„Bill 6 hours ago

China trade war has been paused with a deal in the tea leaves. This will open the door for Apple to both sell better in the Chinese market and eliminate the tariffs possibility on their products made in China. This is the best new for Apple and the Analysts should see it. I see a big jump for AAPL in the coming days and more before the new year.

Reply Replies (2)“

Příspěvky byly kopírovány do textového souboru, včetně přezdivek uživatelů, jejich avatarů a dalších podobných prvků, které nejsou pro analýzu relevantní. Proto bylo nutné textové soubory od těchto prvků očistit. K tomu posloužily textové funkce programů Microsoft Word a Excel. Do analýzy textu tedy vstupoval po čištění komentář v následujícím formátu:

„02.12.2018; China trade war has been paused with a deal in the tea leaves. This will open the door for Apple to both sell better in the Chinese market and eliminate the tariffs possibility on their products made in China. This is the best new for Apple and the Analysts should see it. I see a big jump for AAPL in the coming days and more before the new year.“

2.2.Měření sentimentu

Komentář z předchozí kapitoly byl spolu s ostatními za celé období importován do software STATISTICA. Za pomoci nástroje Text Mining byly potom záznamy příslušící k jednotlivým akciovým titulům porovnány s výše popsanými slovníky od autorů Bing Liu, Elaine Henry a dvojice Tim Loughran a Bill McDonald. Výše zmíněný příspěvek obsahoval slova, která se vyskytla i v jednotlivých slovnících. Jejich počet je vidět v následující tabulce 1.

Tabulka 1: Příklad analýzy sentimentu na vybraném komentáři

Bing Liu	počet
best	1
better	1

Henry	počet
Best	1
better	1
More	1

Loughran, McDonald	počet
better	1

Zdroj: Vlastní zpracování

Po aplikaci textové analýzy v a získání pozitivních a negativních slov ke každému akciovému titulu dle jednotlivých slovníků, bylo třeba konečně naměřený sentiment určitým způsobem kvantifikovat. Pro každý příspěvek byl vyjádřen počet obsažených pozitivních a negativních výrazů. Každý pozitivní výraz zvýšil naměřené skóre o 1, negativní výraz byl zapsán se znaménkem – a skóre o 1 snížil. Součtem těchto hodnot vzniklo tedy absolutní skóre sentimentu, ze kterého bylo zřejmé, zda v konkrétním příspěvku převládá negativní či pozitivní sentiment. Jelikož ale počet výrazů v rámci příspěvku byl různý, bylo nutné vyjádřit toto skóre relativním poměrem. Výhodou je nezávislost na počtu hodnot ve výběru. Matematicky lze zapsat následovně:

$$\sum_{i=1}^k \frac{n_i}{n} = 1, \quad (9)$$

kde n_i je absolutní skóre, zde tedy rozdíl počtu pozitivní a negativních slov, n je rozsah souboru, zde tedy počet pozitivních i negativních slov dohromady a k je počet obměn, tedy jednotlivých analyzovaných komentářů.

Výstupem byly hodnoty v intervalu $\langle -1;1 \rangle$, kde:

-1 = čistě negativní sentiment příspěvku,

0 = neutrální sentiment příspěvku,

1 = čistě pozitivní sentiment příspěvku.

Po vypočtení relativního skóre sentimentu pro každý příspěvek byly hodnoty agregovány v rámci každého jednoho dne pomocí matematického průměru. Takto vypočtené hodnoty podle všech 3 slovníků byly přiřazeny k finančním údajům za měřené období. Tím tedy vznikly tři číselné řady, resp. proměnné pro další analýzu.

Do měřeného období vstoupilo několik víkendů a státních svátků, ve kterých diskuze na sociálních sítích sice probíhala, ale na finančních trzích se aktivně neobchodovalo, nejsou známy finanční údaje, proto byla data za tyto dny vynechána.

2.2.1. Analýza komentářů

Následuje přehled nejčastějších slov, která byla jednotlivými slovníky identifikována (tabulka 2 až 5). Jednomu akciovému titulu přísluší trojice tabulek pod sebou, kde lze porovnat výstup v závislosti na použitém slovníku. Identifikovaná slova pak budou mít stěžejní vliv na výpočet celkového sentimentu. Sloupec počet označuje nikoli absolutní výskyt daného slova, ale počet uživatelských příspěvků, ve kterých se dané slovo objevilo alespoň jednou.

Tabulka 2: Nejčastější slova v diskuzi AAPL, AMZN

AAPL Loughran, McDonald			
pozitivní	počet	negativní	počet
good	1268	bad	571
great	569	cut	473
better	549	lose	468

AMZN Loughran, McDonald			
pozitivní	počet	negativní	počet
good	599	loss	366
great	279	lose	294
better	213	bad	242

AAPL Henry			
pozitivní	počet	negativní	počet
good	1268	drop	770
high	623	low	652
better	549	fall	591

AMZN Henry			
pozitivní	počet	negativní	počet
good	599	drop	337
high	314	low	284
growth	228	fall	259

AAPL Bing Liu			
pozitivní	počet	negativní	počet
trump	2952	fall	591
like	2023	bad	571
good	1268	crash	508

AMZN Bing Liu			
pozitivní	počet	negativní	počet
like	905	loss	366
trump	705	lose	294
good	599	fall	259

Zdroj: Vlastní zpracování

Datový soubor obsahující příspěvky z diskuze k akciovému titulu Apple Inc. Byl jednoznačně největší ze všech pozorovaných. Tabulka o tom jasně vypovídá, počty příspěvků se slovy obsahující sentiment se zde pohybují i v tisících, u většiny ostatních titulů je to spíše v řádu stovek. Slovo „good“ se jako jediné objevilo v analýze na základě všech použitých slovníků, ostatní slova jsou ale v závislosti na použitém slovníku dosti odlišná, což potvrzuje tvrzení uvedené výše, o naprosto klíčové roli použitého slovníku. Slovo „good“ bylo jedním z nejčastěji použitých slov k vyjádření pozitivního smýšlení u všech osmi titulů analyzovaných podle všech tří slovníků. Při porovnání slov s negativním sentimentem je ale trojice nejčastěji použitých slov rozmanitější.

Tabulka 3: Nejčastější slova v diskuzi BABA, BAC

BABA Loughran, McDonald			
pozitivní	počet	negativní	počet
good	444	bad	127
great	157	break	94
strong	135	miss	80

BAC Loughran, McDonald			
pozitivní	počet	negativní	počet
good	535	bad	245
great	353	lie	240
win	191	investing	156

BABA Henry			
pozitivní	počet	negativní	počet
good	444	low	192
high	203	drop	135
growth	193	lower	105

BAC Henry			
pozitivní	počet	negativní	počet
good	535	drop	225
best	179	low	163
better	174	worst	135

BABA Bing Liu			
pozitivní	počet	negativní	počet
good	444	bad	127
like	404	break	94
trump	401	fall	84

BAC Bing Liu			
pozitivní	počet	negativní	počet
trump	2768	bad	245
like	850	lie	240
good	535	hate	161

Zdroj: Vlastní zpracování

Druhá tabulka shrnuje údaje společností Alibaba Group Holding Limited a Bank of America Corporation. Jediným slovem, které se zde vyskytovalo ve více než tisíci příspěvcích, je slovo „trump“. Toto slovo figuruje ve všech analýzách podle slovníku autora Bing Liu, v ostatních slovnících zastoupeno není. Projevuje se zde jeden z výše zmíněných nedostatků analýzy. Toto slovo může v anglickém jazyce znamenat trumf, což je entita, která nese pozitivní sentiment. Také je to ale označení jiné entity, a to příjemní současného prezidenta Spojených států amerických. Všechna rozhodnutí nebo vyjádření této celosvětově významné osobnosti mají mimo jiné dopad i na finanční trhy. Z toho důvodu je často zmiňován i v internetových diskuzích a lze proto předpokládat, že značná část identifikovaných příspěvců s tímto slovem označuje právě Donalda Trumpa. V řadách jeho příznivců nese jeho jméno pravděpodobně pozitivní sentiment, mezi jeho odpůrci to ale bude pravděpodobně naopak, případně bude toto jméno vnímáno neutrálně, s tím však analýza sentimentu nedokáže pracovat. Každý výskyt slova „trump“ je zde započten jako ta samá entita s pozitivním sentimentem a může zde tedy docházet k drobnému zkreslení.

Tabulka 4: Nejčastější slova v diskuzi CSCO, FB

CSCO Loughran, McDonald			
pozitivní	počet	negativní	počet
good	371	lie	437
win	243	poor	345
great	209	investig	227

FB Loughran, McDonald			
pozitivní	počet	negativní	počet
good	320	bad	212
great	154	lose	136
better	152	negat	126

CSCO Henry			
pozitivní	počet	negativní	počet
good	371	low	120
leader	122	-	-
best	117	-	-

FB Henry			
pozitivní	počet	negativní	počet
good	320	low	176
better	152	drop	141
high	126	negat	126

CSCO Bing Liu			
pozitivní	počet	negativní	počet
trump	3718	lie	437
like	805	poor	345
good	371	fake	304

FB Bing Liu			
pozitivní	počet	negativní	počet
like	607	bad	212
good	320	fake	159
trump	185	lose	136

Zdroj: Vlastní zpracování

Ve třetí tabulce je shrnutí za společnosti Cisco Systems, Inc. A Facebook, Inc. V případě použití slovníku autora Bing Liu je opět patrný extrémně vysoký výskyt slova „trump“. Další zajímavý jev je pouze jedno identifikované negativní slovo podle slovníku autorky Elaine Henry u titulu CSCO. Počet analyzovaných příspěvků přitom není nižší než u ostatních titulů. Je tedy možné, že ve sledovaném období téměř nikdo v diskuzi nevyjadřoval negativní sentiment, v tom případě by však použití ostatních slovníků mělo vypovídat o tom samém. Pravděpodobnější se tedy zdá vysvětlení, že tento slovník není vhodný pro specifickou diskuzi ke společnosti Cisco Systems, Inc.

Tabulka 5: Nejčastější slova v diskuzi T, WMT

T Loughran, McDonald			
pozitivní	počet	negativní	počet
good	221	cut	121
great	108	bad	87
better	97	lose	84

WMT Loughran, McDonald			
pozitivní	počet	negativní	počet
good	143	crimin	73
great	83	lie	72
win	50	illeg	56

T Henry			
pozitivní	počet	negativní	počet
good	221	low	183
high	104	drop	114
better	97	lower	64

WMT Henry			
pozitivní	počet	negativní	počet
good	143	low	39
better	42	less	33
best	40	least	32

T Bing Liu			
pozitivní	počet	negativní	počet
like	328	debt	167
good	221	bad	87
trump	155	lose	84

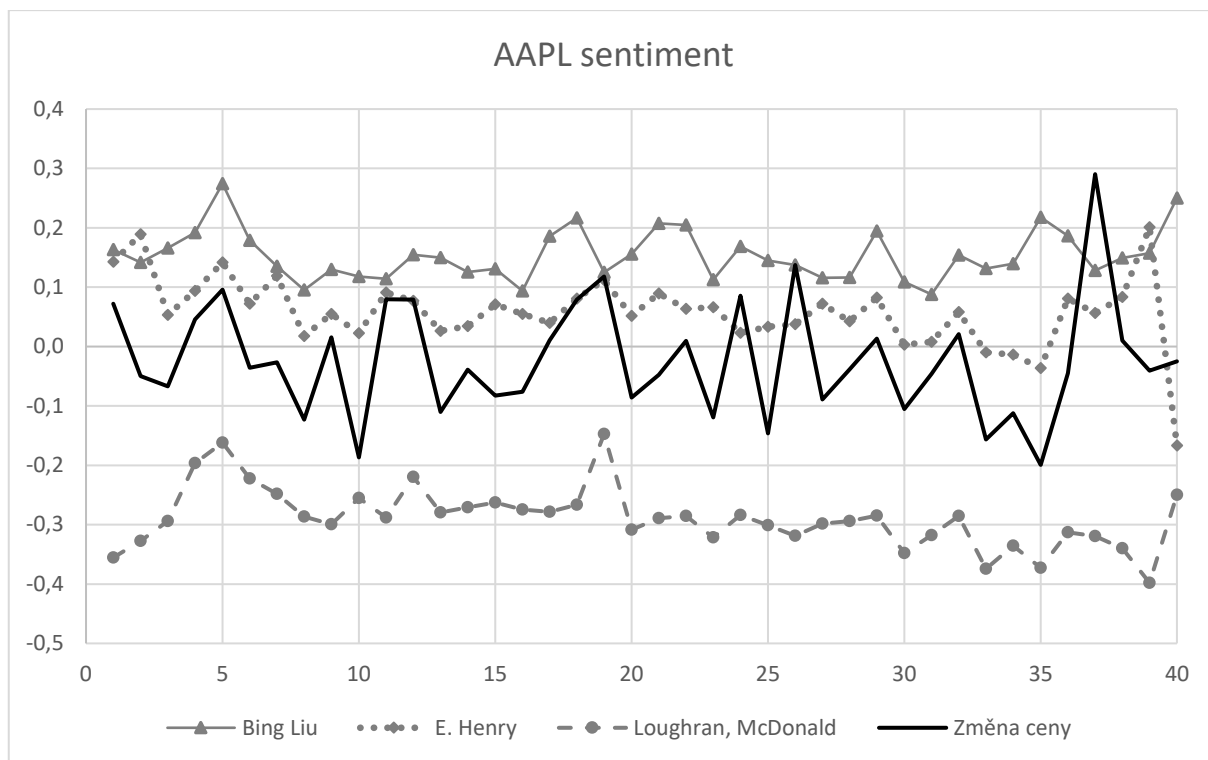
WMT Bing Liu			
pozitivní	počet	negativní	počet
trump	457	lie	72
like	196	crimin	68
good	143	illeg	55

Zdroj: Vlastní zpracování

Čtvrtá tabulka obsahuje údaje za akciové tituly AT&T Inc. a Walmart Inc. U nich obsahovala diskuze méně příspěvků, proto je v některých případech nižší také počet příspěvků se slovy, respektive kořeny slov, nesoucími sentiment. Pro další srovnávání budou však použity relativní četnosti, méně příspěvků tedy nebrání v dalším postupu. Titul WMT je jediný, kde se mezi nejvlivnějšími slovy objevil kořen „*illeg*“, který je obsažen například ve slovech *illegal*, *illegally*. Společnost Walmart byla v období sběru dat kritizována v médiích za údajné ilegální praktiky při přijímání zaměstnanců. Častý výskyt tohoto slova by mohl s touto kauzou souviset. V další kapitole posoudíme, zdali se tato obvinění promítla také do cen akcií.

2.2.2. Vývoj sentimentu

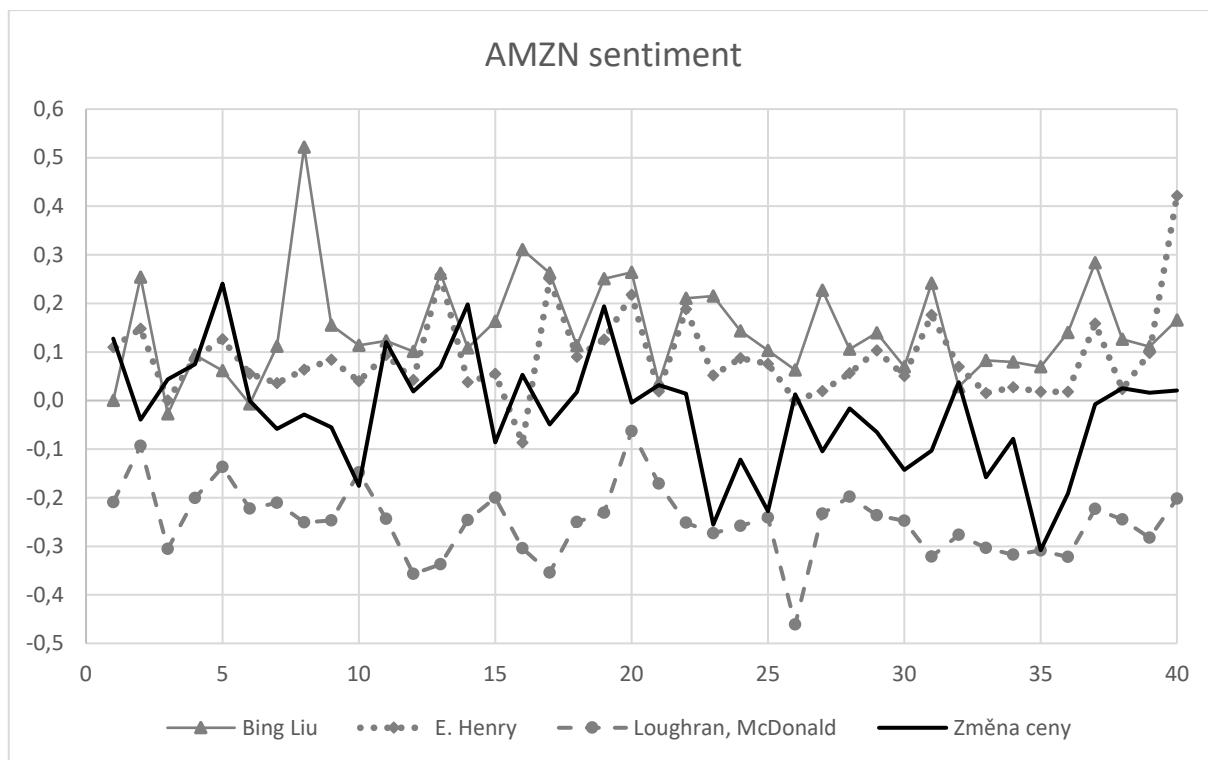
Údaje byly zaneseny do grafů (obrázek 6 až 13), kde lze nyní porovnat relativní vývoj sentimentu dle jednotlivých slovníků a ceny. Křivka popisující ceny v grafech vychází pro lepší přehlednost z desetinásobku procentní změny. Pokud by bylo zachováno skutečné měřítko, na grafu by se přímka ceny zdála v porovnání s křivkami sentimentu téměř konstantní, což by znesnadnilo porovnávání. V názvu grafu je akciový titul, ke kterému se hodnoty vztahují. Osa X znázorňuje 40 obchodních dní, ke kterým byla shromážděna data. Osa Y vyjadřuje relativní úroveň sentimentu a změny ceny ve výše popsaném intervalu.



Obrázek 6: Vývoj sentimentu Apple Inc.

Zdroj: Vlastní zpracování

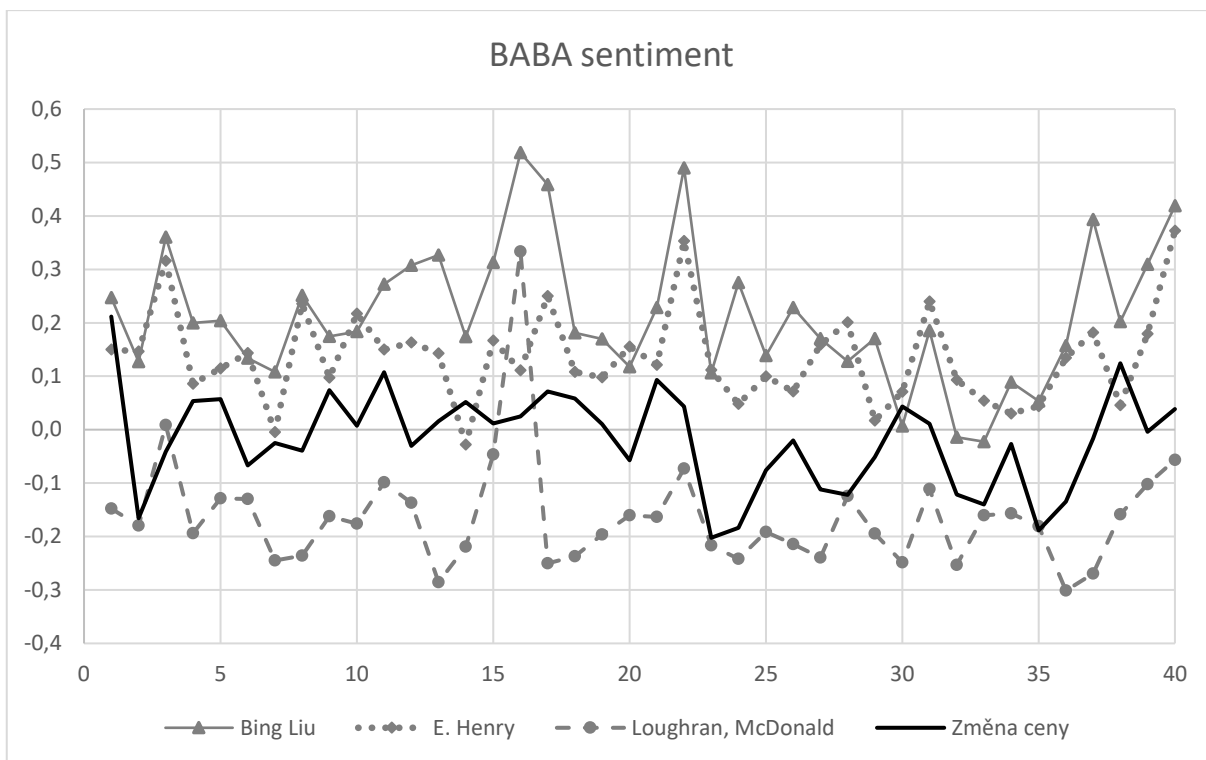
Vývoj sentimentu i ceny u titulu AAPL spolu hned v počátečních několika dnech koresponduje poměrně dobře. Pátý den stoupala shodně cena i sentiment podle všech slovníků. Za zmínku stojí dále 11. a 19. den, kdy se všechny křivky, kromě sentimentu podle Loughrana a McDonalda, sešly na hodnotě 0,1 což lze interpretovat jako že kladný sentiment převýšil záporný o 10 % a zároveň stoupla cena akcií o 1 %. Dále už byl však vývoj různorodější. Sentiment podle Bing Liu se po celé období držel na vyšších hodnotách než ostatní ukazatele, pravděpodobně tedy její slovník zachytil více pozitivních výrazů. Naopak sentiment podle dvojice autorů Loughrana a McDonalda byl po celou dobu na nižší úrovni. Zaznamenal tedy více slov s negativním sentimentem.



Obrázek 7: Vývoj sentimentu Amazon.com, Inc.

Zdroj: Vlastní zpracování

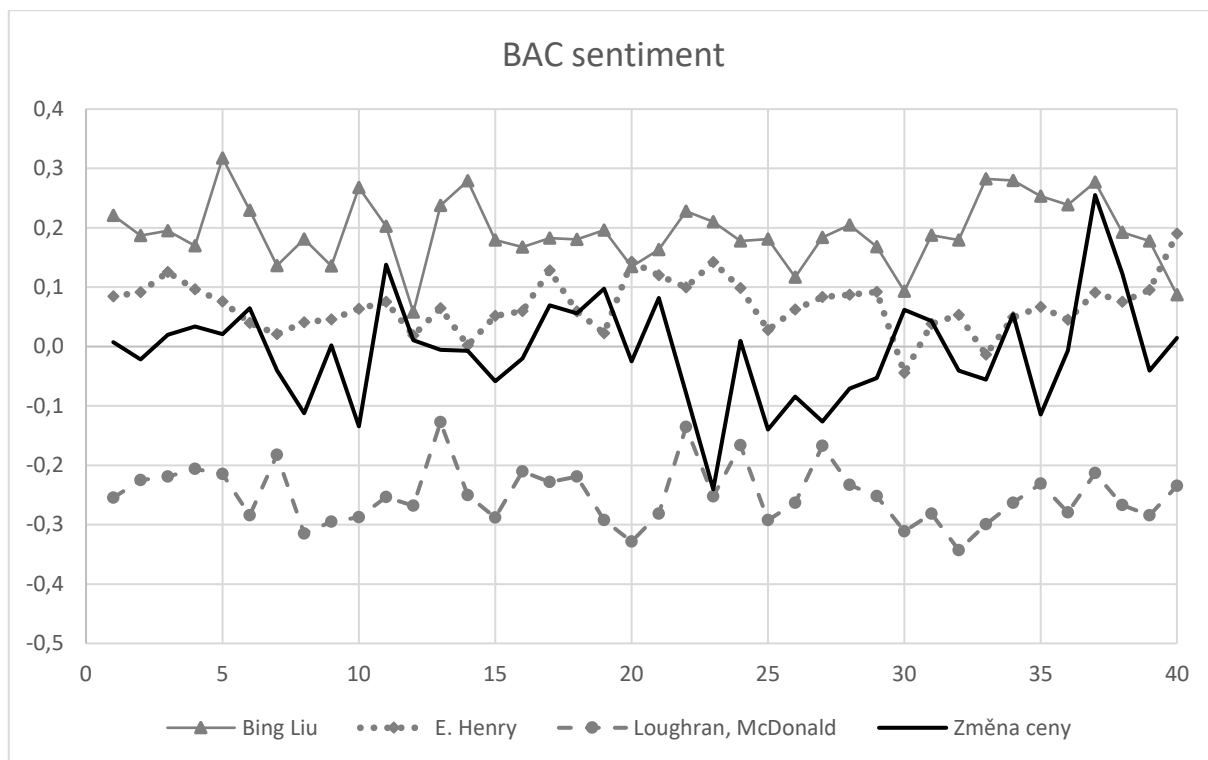
U titulu AMZN jsou si navzájem jednotlivé křivky trochu bližší než v předchozím případě. V grafu ale na první pohled nejsou viditelná místa se stejným trendem ve vývoji ceny i sentimentu. Cena nejvíce stoupla pátý den, nejvíce klesala v třicátém pátém dni a naměřený sentiment v těchto dnech žádných extrémů nedosahoval. Mezi jedenáctým a dvacátým druhým dnem cena převážně rostla, mezi dvacátým druhým a třicátým sedmým dnem zase převážně klesala. Z posouzení grafu se nezdá, že by některé měření sentimentu mělo stejný vývoj, jelikož se křivky stabilně drží buď v kladných anebo záporných hodnotách. Vztahy budou důkladněji prozkoumány v další části práce.



Obrázek 8: Vývoj sentimentu Alibaba Group Holding Limited.

Zdroj: Vlastní zpracování

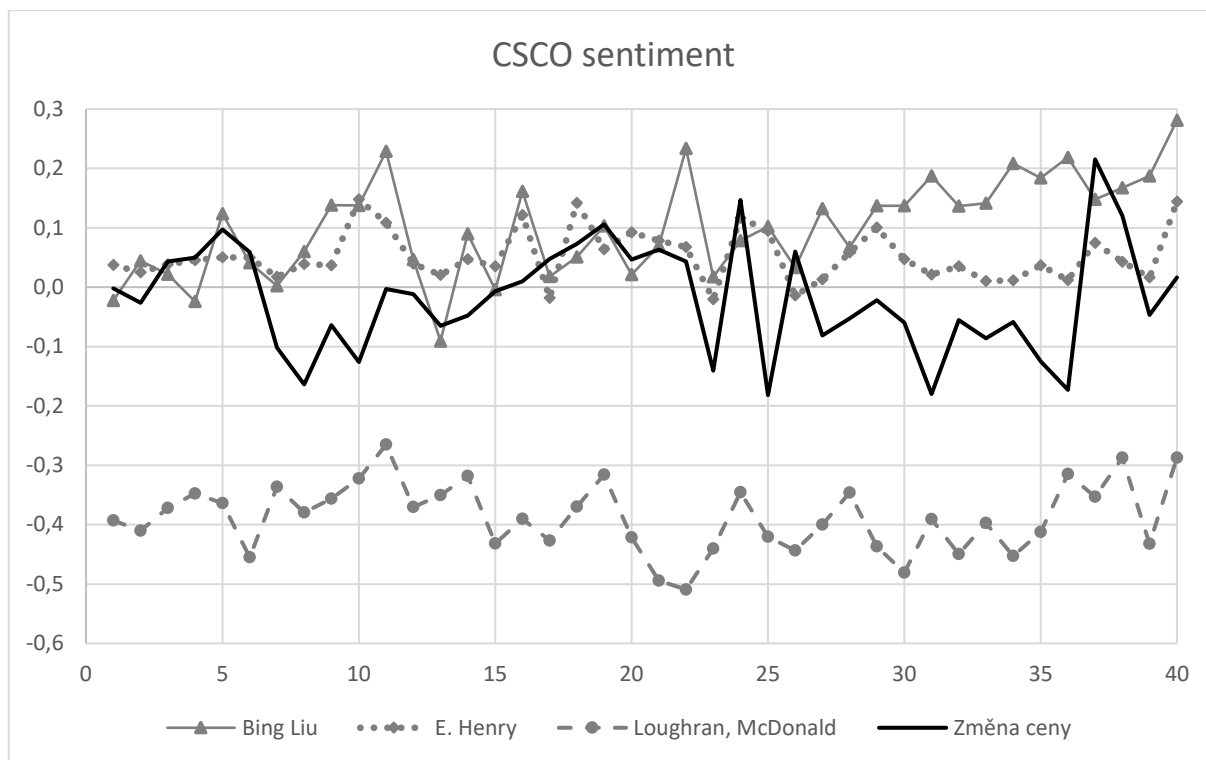
Pro titul BABA také platí, že vypočtené hodnoty jsou nedaleko od sebe. Šestnáctý den zaznamenala analýza podle slovníků Bing Liu a Loughrana McDonald strmě nárůst pozitivního sentimentu, cena se však nepohnula o více než 1 %. Zajímavým momentem byl také 21. a 22. den, kdy shodně stoupaly všechny sledované křivky. Podobný vývoj nastal i na konci sledovaného období. Sentiment podle všech slovníku reflektoval vývoj ceny, dokonce se zdá, jako by cena s mírným zpožděním kopírovala sentiment podle Bing Liu. Jelikož smyslem této práce není pouze sentiment kvantifikovat, ale především zjistit, zda se dá využít k predikci, mohlo by toto období mít velký význam v další analýze.



Obrázek 9: Vývoj sentimentu Alibaba Group Holding Limited.

Zdroj: Vlastní zpracování

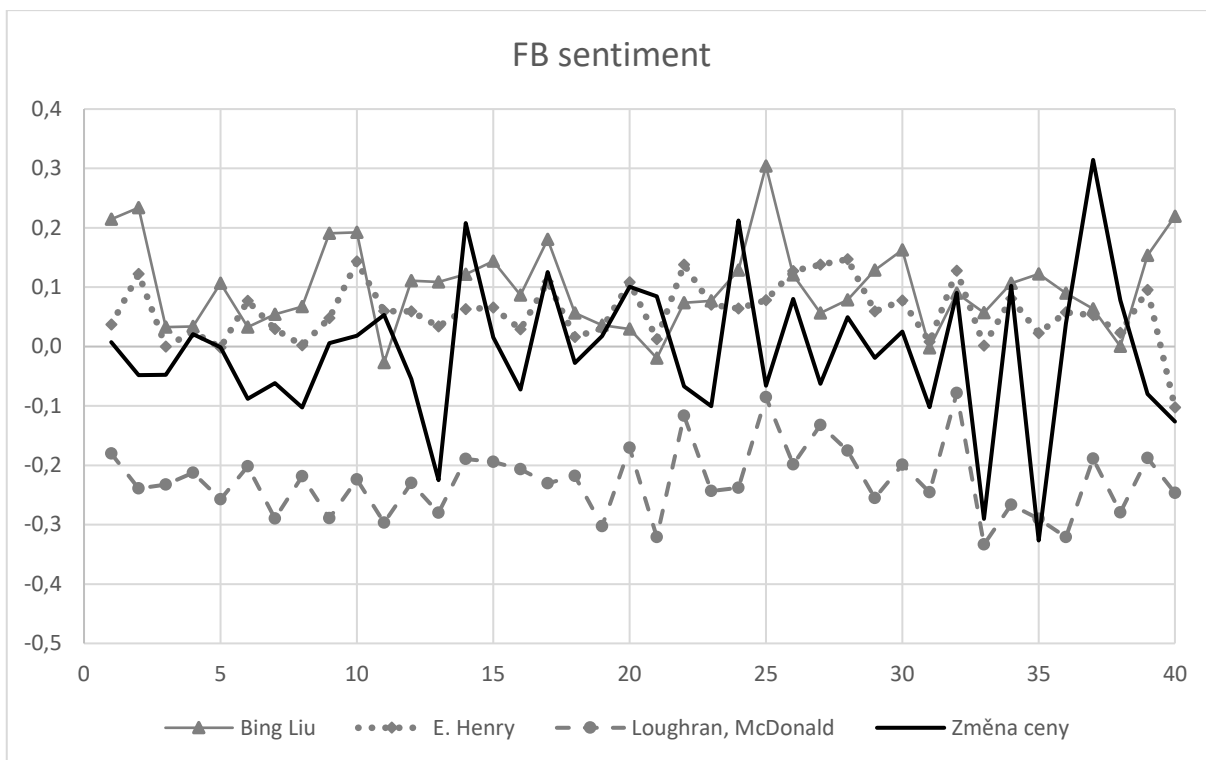
Vývoj ceny titulu BAC vytvořil mezi pátým a dvanáctým dnem podobnou formaci, jako sentiment podle slovníku Bing Liu. Mezi dvacátým a dvacátým pátým dnem se vývoj zase podobá křivce podle slovníku Loughrana a McDonalda. Výrazná změna v ceně nastala třicátý sedmý den, i v sentimentu podle všech tří autorů je patrná změna směrem vzhůru, nicméně není zdaleka tolik výrazná. Zajímavý je i úsek od prvního dne k pátému dni, kdy cena i sentimenty vykazují podobně nízkou volatilitu.



Obrázek 10: Vývoj sentimentu Cisco Systems, Inc.

Zdroj: Vlastní zpracování

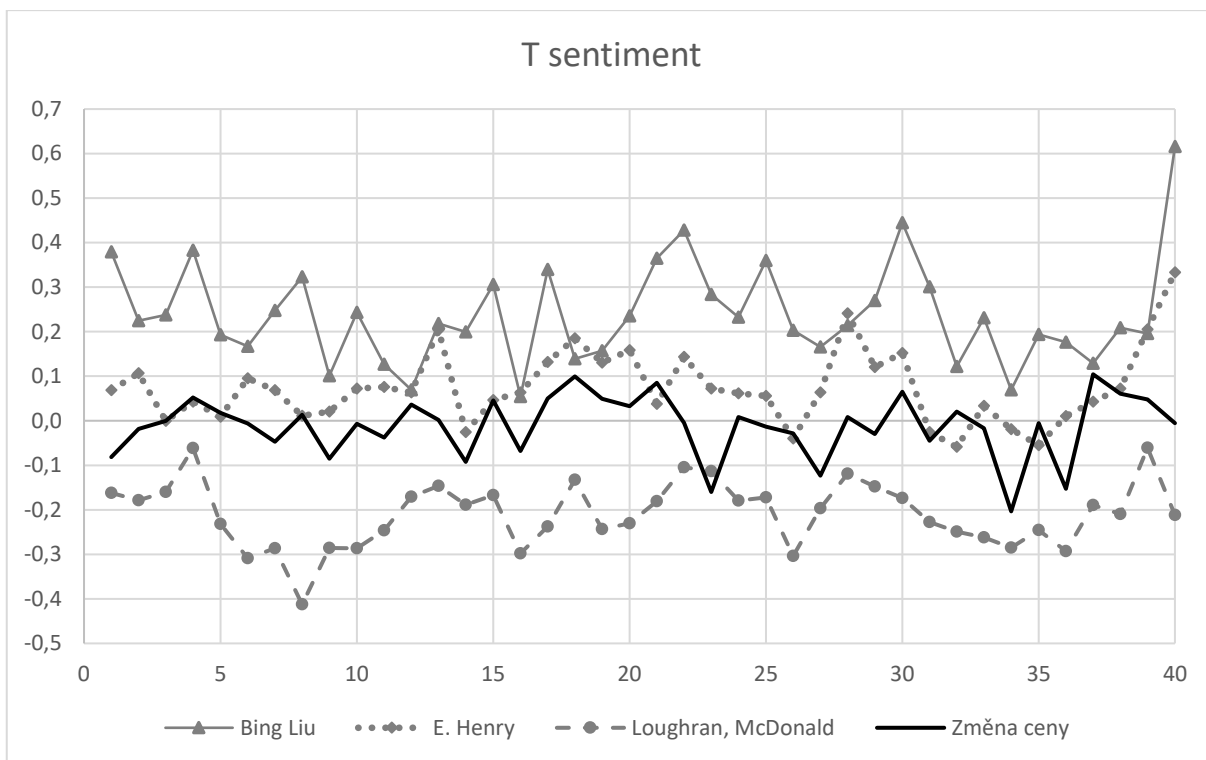
Výsledný sentiment podle Loughrana a McDonalda je u titulu CSCO výrazně na nižší úrovni než ostatní křivky. V analýze tedy zřejmě za celé období zachytil výrazně vyšší počet negativních slov nežli pozitivních. Například jedenáctý, devatenáctý a dvacátý čtvrtý den však vytvořil vrcholy shodně jako vývoj ceny. Mezi třináctým a dvacátým druhým dnem vývoj ceny překrývá vývoj sentimentu podle Bing Liu a E. Henry, byť sentiment vykazuje v obou případech větší volatilitu. Třicátý sedmý den však došlo k extrémnímu nárůstu ceny. Ten se na žádné křivce sentimentu neprojevil tolik výrazně.



Obrázek 11: Vývoj sentimentu Facebook, Inc.

Zdroj: Vlastní zpracování

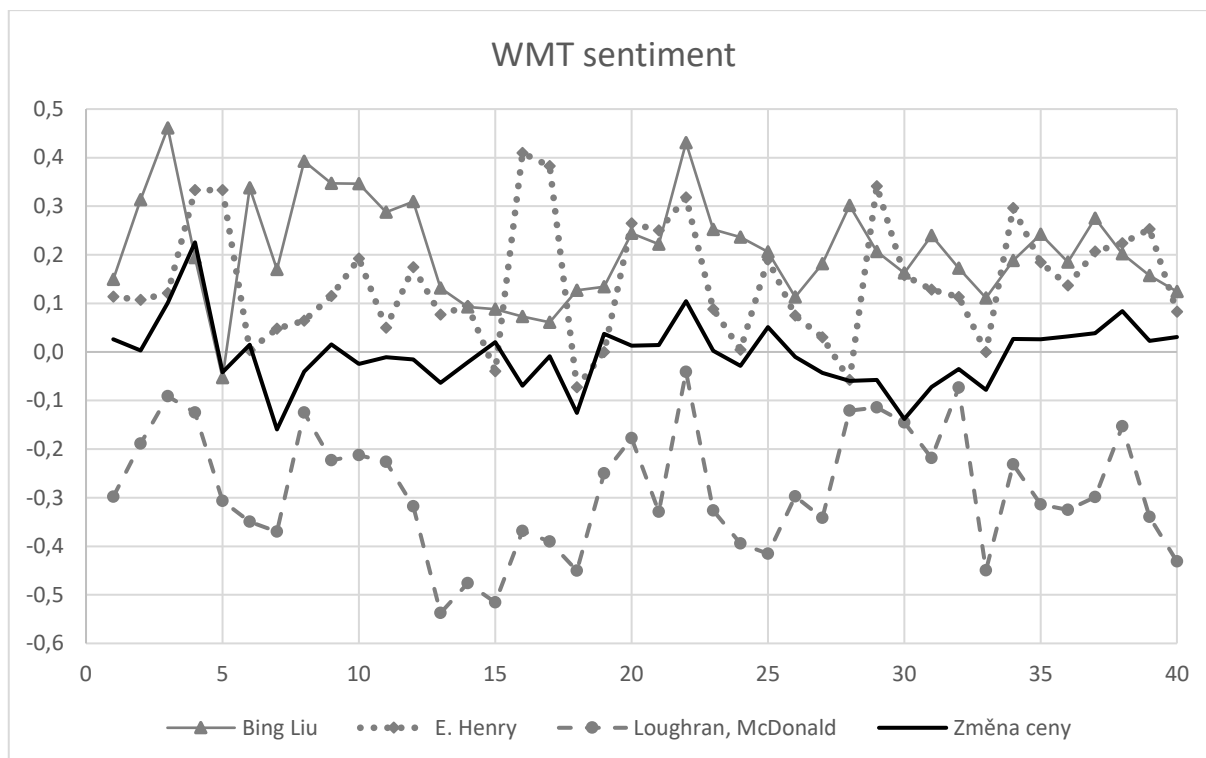
Titul FB dosahoval ve vývoji ceny největší volatility. I na grafech sentimentu jsou patrné výrazné výběžky, někdy však cenu předbíhají, někdy kopírují, někdy se opožďují. Aby se údaje daly použít pro predikci, je potřeba nalézt vztah s určitou konzistencí. Další část práce se bude zabývat hledáním právě takového vztahu.



Obrázek 12: Vývoj sentimentu AT&T Inc.

Zdroj: Vlastní zpracování

Cena titulu T nevykazovala ve sledovaném období příliš velkou volatilitu. Sentimenty podle slovníku různých autorů se vyvíjely podobně, například mezi dvacátým osmým a třicátým pátým dnem. Jinde však na první pohled není zřejmý vztah s vývojem ceny. Osmý den byl naměřen výrazný pokles sentimentu podle autorů Loughrana a McDonalda, vývoj ceny a sentimentu podle E. Henry byl naprosto neutrální a sentiment podle Bing Liu šel dokonce opačným směrem. V tomto bodu se tedy jednotlivé přístupy naprosto rozcházejí. Ve čtyřicátý den sentiment podle Bing Liu a E. Henry výrazně rostl, cena se však vyvíjela jiným směrem.



Obrázek 13: Vývoj sentimentu Walmart Inc.

Zdroj: Vlastní zpracování

Křivka vývoje ceny titulu WMT má opět výrazně nižší volatilitu než naměřený sentiment. Maxima dosáhla čtvrtý den, sentiment podle slovníku Bing Liu i Loughrana a McDonalda ji nepatrně předběhl, sentiment podle slovníku E. Henry vývoj kopíroval. Další podobnost lze vysledovat mezi devatenáctým a dvacátým čtvrtým dnem, kdy všechny křivky vytvořily podobnou formaci. V předešlé kapitole bylo zjištěno, že kořen slova „ilegall“ byl jedním z nejčastějších nalezených v příspěvcích. Volatilita sentimentu svědčí o bouřlivé diskuzi ve sledovaném období, vývoj ceny však tak dramatické výkyvy neprodělal.

2.3. Predikce změn ceny akcií na základě analýzy sentimentu

V této kapitole půjde o zjištění vlivu naměřených hodnot sentimentu na vývoj ceny v odpovídajících dnech. Prvním krokem je definování proměnných.

2.3.1. Popis proměnných

První proměnná „kod spol“ je nominální proměnná. Nominální proměnné často slouží pro identifikaci objektů anebo jsou v některých úlohách podkladem pro jejich třídění [13]. Zde proměnná identifikuje danou společnost. Je složena z osmi variant podle osmi akciových titulů

do regresní analýzy vstupuje jako nezávislá proměnná neboli regresor, a může to být důležitý determinant.

Další nezávisle proměnné jsou vypočtená skóre získaná analýzou sentimentu podle různých autorů slovníků. Nazvány jsou „**Bing Liu, Henry a Loughran, McDonald**“, nabývají hodnot v intervalu $\langle -1; 1 \rangle$, kde -1 vypovídá o 100 % negativním sentimentu v daném dnu, hodnota 1 značí 100% pozitivní sentiment v daném dnu.

Třetí skupinou regresorů jsou kvantitativní proměnné vypočtené na základě ceny v jednotlivých dnech, tedy vybrané ukazatele technické analýzy. Vzorce jejich výpočtu jsou popsány v kapitole 1.3.4. Indikátory technické analýzy. Proměnné jsou označeny názvy:

- **SMA 3** (Simple Moving Average třídenní)
- **SMA 5** (Simple Moving Average pětidenní)
- **SMA 10** (Simple Moving Average desetidenní)
- **EMA 5** (Exponential Moving Average pětidenní)
- **RSI 9** (Relative Strength Index devítidenní)
- **Stochastic** (Stochastický oscilátor čtrnáctidenní)

A konečně závislá proměnná „**vynos**“, udává výnos následujícího dne. Byla určena následovně:

$$vynos = \frac{\text{konečná cena} - \text{počáteční cena}}{\text{počáteční cena}} \quad (10)$$

Prvním krokem je však zjištění korelace mezi proměnnými. Použit byl Pearsonův korelační koeficient, který se počítá pomocí směrodatných odchylek obou proměnných a jejich kovariance (kovariance = míra vzájemné vazby mezi veličinami) [4]:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} = \frac{S_{xy}}{S_x S_y} \quad (11)$$

Na základě testu významnosti lze říci, zda mezi náhodnými veličinami X a Y existuje korelační vztah: $H_0: \rho = 0$ $H_A: \rho \neq 0$

Výstup z programu STATISTICA shrnuje zjištěné Pearsonovy koeficienty korelace (tabulka 6).

Tabulka 6: Korelace proměnných vstupujících do regrese

Korelace. Označ. korelace jsou významné na hlad. $p < ,05000$ $N=216$ (Celé případy vynechány u ChD)										
Proměnná	Bing Liu	Henry	Loughran, McDonald	SMA 3	SMA 5	SMA 10	EMA 5	RSI 9	Stochastic	vynos
Bing Liu	1,000	0,277	0,408	-0,093	-0,095	-0,095	-0,093	-0,078	0,064	0,117
Henry	0,277	1,000	0,311	0,023	0,018	0,016	0,022	0,075	0,223	0,155
Loughran, McDonald	0,408	0,311	1,000	0,028	0,026	0,023	0,028	0,027	0,223	0,131
SMA 3	-0,093	0,023	0,028	1,000	1,000	0,999	1,000	0,087	0,143	-0,010
SMA 5	-0,095	0,018	0,026	1,000	1,000	0,999	1,000	0,084	0,136	-0,012
SMA 10	-0,095	0,016	0,023	0,999	0,999	1,000	0,999	0,073	0,124	-0,011
EMA 5	-0,093	0,022	0,028	1,000	1,000	0,999	1,000	0,086	0,141	-0,010
RSI 9	-0,078	0,075	0,027	0,087	0,084	0,073	0,086	1,000	0,693	-0,034
Stochastic	0,064	0,223	0,223	0,143	0,136	0,124	0,141	0,693	1,000	0,325
vynos	0,117	0,155	0,131	-0,010	-0,012	-0,011	-0,010	-0,034	0,325	1,000

Zdroj: Vlastní zpracování

V případě tohoto výběru tedy výsledné korelační koeficienty jsou nenulové, platí tedy $H_A: \rho \neq 0$. Řada z nich je však nevýznamná na hladině významnosti $\alpha = 0,05$. Velmi silné vztahy závislosti jsou však mezi proměnnými SMA 3, SMA 5, SMA 10, EMA 5. Takto vysoké koeficienty korelace by v regresním modelu způsobily chybu. V dalším kroku je třeba aplikovat nástroj, který umožní tyto proměnné nahradit jinými, nekorelovanými proměnnými.

2.3.2. Faktorová analýza

Faktorová analýza byla původně užívána v psychologii. Poprvé ji zde použil Charles Spearman pro testy inteligence. Její užití je vhodné v případě, kdy máme v analýze větší množství proměnných, které ale vypovídají o tom samém, resp. měří stejnou veličinu, přestože mohou být sestaveny jiným způsobem. Často se stává, že počet pozorovaných znaků je redundantní z důvodu silných korelací mezi znaky. V takovém případě je možné téměř celou informaci, resp. „variabilitu“, která se v těchto znacích vyskytuje, zachytit výrazně menším počtem přímo nepozorovatelných proměnných [10].

Při ověřování předpokladů regresní analýzy se narazilo na vysokou vzájemnou korelaci především u ukazatelů založených na klouzavém průměru. Regresní analýza by tím byla zkreslená. Faktorová analýza nám ale poslouží k identifikování vztahů (faktorů), které stojí v pozadí vypočtených hodnot a tím sjednotí a zredukuje počet nezávislých proměnných. Počet faktorů bývá často stanoven na základě znalosti vzorku, pokud jsou si některé proměnné podobné. To je v tomto případě zřejmé, ve faktorové analýze půjde o zastoupení proměnných SMA 3, SMA 5, SMA 10 a EMA 5 jedním faktorem. Pokud by bylo nutné vysvětlit více faktory více proměnných, jejich počet by bylo nutné odhadnout na základě znalosti dat nebo například grafickou metodou. Grafickou metodou pro stanovení počtu faktorů je sutinový graf, známý též jako scree plot. Jedná se vlastně o graf vlastních čísel (rozptylů) všech faktorů. Křivka grafu je klesající a mez oddělující vhodný počet faktorů se nachází tam, kde vidíme nejvyšší pokles vlastních čísel mezi dvěma faktory [37].

Cílem faktorové analýzy je obecně výpočet matice F . Tato matice se nazývá faktorová matice. K odhadu matice F existuje řada metod, jako např. modifikovaná metoda hlavních komponent, metoda maximální věrohodnosti, Jöreskogova metoda, centroidová metoda [21]. Využitím ortogonální rotace lze matici faktorových zátěží interpretovat jako regresní koeficienty závislosti proměnných na faktorech, ale také zároveň jako korelační koeficienty vztahu mezi proměnnými a faktory. Při šikmé rotaci lze získat matici regresních či korelačních koeficientů a také matici, která ukazuje korelaci mezi výslednými faktory [13].

Bez ohledu na metodu extrakce faktorů existuje nekonečně mnoho faktorových řešení. Znamená to, že ke každému odhadu faktorových zátěží existuje nekonečně mnoho dalších alternativ, které vystihují data stejně dobře [38]. Při výběru rotace i počtu faktorů tak hraje velkou roli subjektivita. Pro účely této práce však stačilo určit pouze jeden faktor, extrakce byla provedena metodou hlavních komponent. Díky tomu jsou koeficienty korelace mezi proměnnými nižší (tabulka 7).

Tabulka 7: Korelace proměnných se zastoupením faktorem

Proměnná	Korelace; Označ. korelace jsou významné na hlad. $p < ,05000$; $N=216$ (Celé případy vynechány u ChD)						
	Bing Liu	Henry	Loughran, McDonald	RSI 9	Stochastic	Faktor MA	Výnos
Bing Liu	1,000000	0,276758	0,407705	-0,078265	0,063901	0,094224	0,116941
Henry	0,276758	1,000000	0,310797	0,074586	0,223383	-0,019602	0,155424
Loughran, McDonald	0,407705	0,310797	1,000000	0,026714	0,222582	-0,026053	0,131429
RSI 9	-0,078265	0,074586	0,026714	1,000000	0,692953	-0,082717	-0,034490
Stochastic	0,063901	0,223383	0,222582	0,692953	1,000000	-0,135705	0,324987
Faktor MA	0,094224	-0,019602	-0,026053	-0,082717	-0,135705	1,000000	0,010831
Výnos	0,116941	0,155424	0,131429	-0,034490	0,324987	0,010831	1,000000

Zdroj: Vlastní zpracování

Počet nezávislých proměnných se tedy snížil. V tomto formátu nyní lze pokračovat regresní analýzou.

2.3.3. Regresní analýza

Jednou z nejpoužívanějších metod pro odhalení takové závislosti je regresní analýza. Lineární regresní model, který popisuje vztah závislosti mezi vysvětlovanou proměnnou Y a vysvětlujícími proměnnými X_1, X_2, \dots, X_k , lze zapsat následovně [15]:

$$Y = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + u, \quad (12)$$

kde u je náhodná složka,

β_j – j -tý regresní koeficient či parametr ($j = 1, 2, \dots, k$).

X_1 lze specifikovat jako zvláštní (umělou) proměnnou, která nabývá ve všech pozorováních hodnoty rovné jedné, takže vztah lze psát jako [15]:

$$Y = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k + u, \quad (13)$$

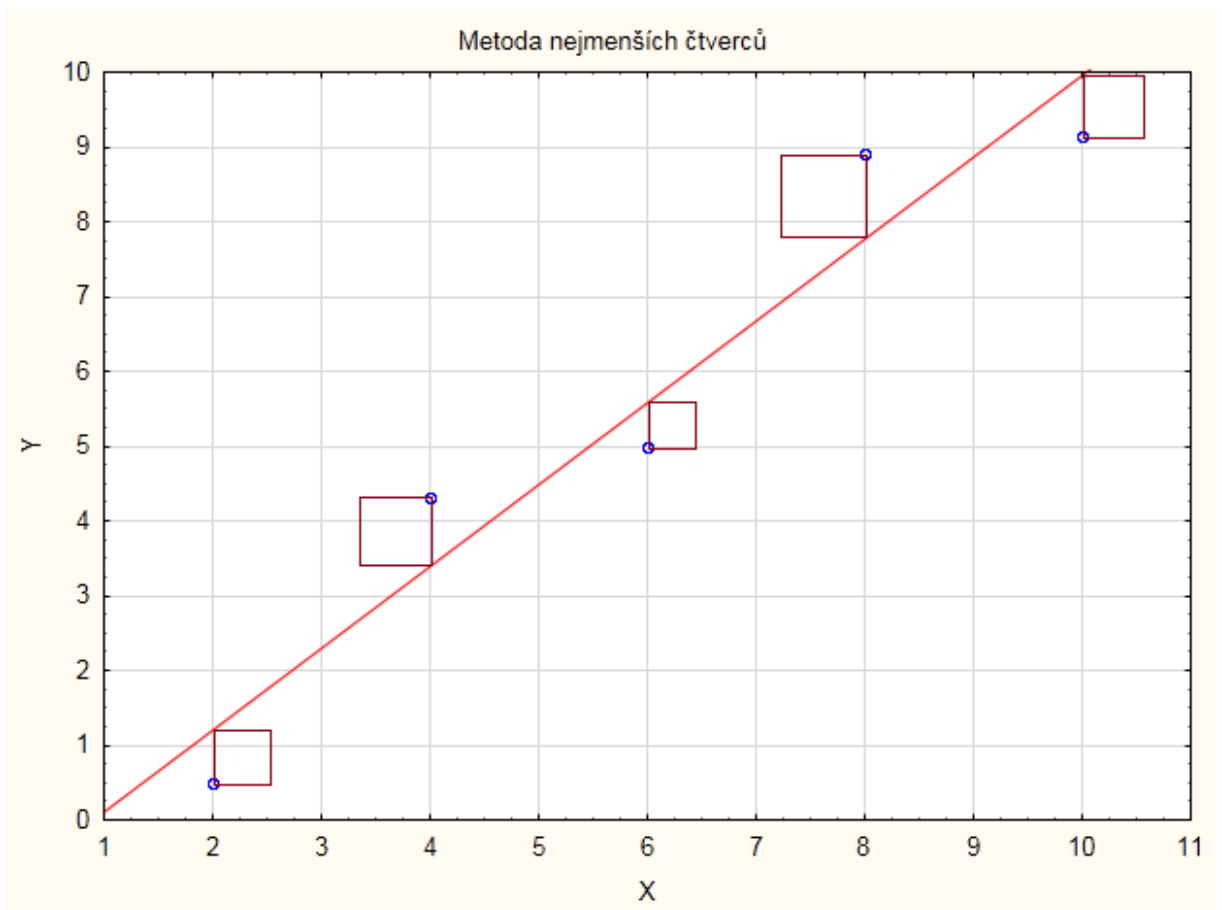
přičemž β_1 je tzv. absolutní člen nebo úroňová konstanta .

Jestliže pro náhodnou složku platí, že $E(u) = 0$, pak očekávanou hodnotou Y jako funkci daných hodnot vysvětlujících proměnných můžeme vyjádřit pomocí deterministického vztahu [15]:

$$E(Y) = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k, \quad (14)$$

který se také nazývá regresní funkcí základního souboru. Regresní koeficienty $\beta_2, \beta_3, \dots, \beta_k$ měří změnu $E(Y)$ odpovídající jednotkové změně libovolné jedné vysvětlující proměnné, přičemž ostatní vysvětlující proměnné zůstávají konstantní (*ceteris paribus*) [15].

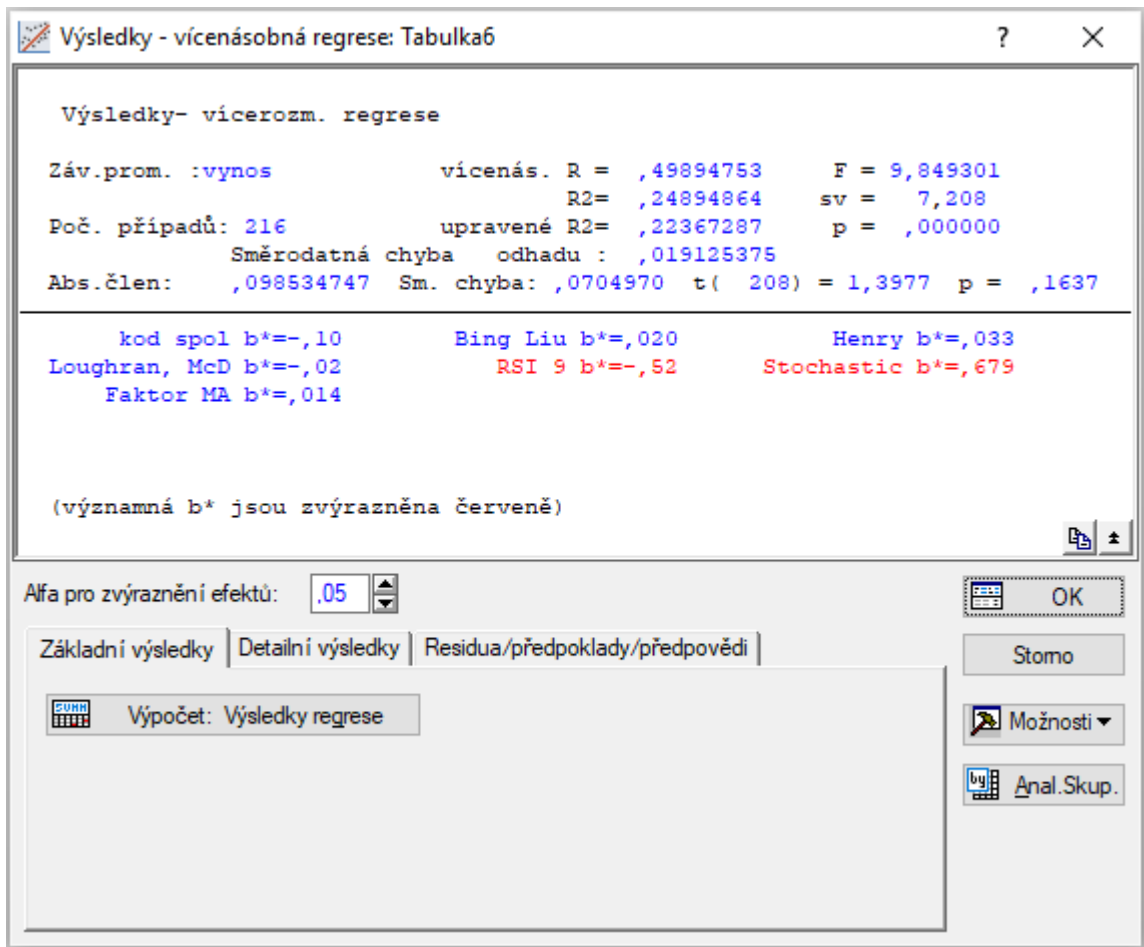
Při odhadu koeficientů lze použít řadu metod, nejběžnější je **Metoda nejmenších čtverců**. Ve statistice se objevuje v různých modifikacích, vždy jde ale o minimalizování součtu druhých mocnin odchylek. Graficky jsou tyto druhé mocniny znázorněny čtverci (obrázek 14).



Obrázek 14: Metoda nejmenších čtverců

Zdroj: [1]

Výsledky vícerozměrné regrese jsou prezentované na obrázku 15. V něm jsou uvedeny jak hodnoty regresních koeficientů pro jednotlivé nezávislé proměnné, tak kvalita regresního modelu a jeho statistická významnost.



Obrázek 15: Regresní analýza v software STATISTICA

Zdroj: Vlastní zpracování

Hodnota R^2 je koeficient determinace. Popisuje, jaký podíl celkové variability v závisle proměnné se podařilo vysvětlit regresním modelem. Zde tedy 24,89 %. Z této hodnoty je patrné, že na vývoj ceny akcií působí mnoho dalších vlivů, které nebyly do tohoto regresního modelu zahrnuty.

Upravené R^2 má podobný význam jako koeficient determinace, bere však také do úvahy počet regresorů zahrnutých v modelu. Mělo by se tedy jednat o přesnější ukazatel, hodnota je zde trochu nižší, 22,37 %.

F/sv/p. Hodnota F , sv a odpovídající hladina p vystupují v roli celkového F -testu vztahů mezi závislou proměnnou a množinou nezávislých proměnných [3]. Zde má následující význam: F

= regresní průměr čtverců / reziduální průměr čtverců. Na základě této hodnoty by se dala porovnat vhodnost různých modelů.

Směrodatná chyba odhadu představuje míru rozptýlení pozorovaných hodnot okolo regresní přímky.

Abs. člen. je pole, které obsahuje odhad b_0 absolutního členu regresní rovnice.

Sm. chyba. je směrodatná chyba absolutního členu.

$t(sv)$ a p . Hodnota t a odpovídající p se používají pro test hypotézy, že absolutní člen je roven 0.

b^* koeficienty jednotlivých nezávisle proměnných jsou zobrazeny ve spodní části souhrnného okna. Tyto koeficienty však neodpovídají odhadům parametrů b_0 a b_1 z uvažovaného regresního modelu. Jedná se o odhady parametrů z jiného modelu, ve kterém by byly nejdříve všechny regresory standardizovány tak, aby jejich průměr byl roven 0 a směrodatná odchylka byla rovna 1. b^* je pak odhad tohoto modelu. Koeficienty b^* proto umožňují porovnat relativní vliv jednotlivých regresorů na závisle proměnnou [3].

2.3.4. Vliv technických indikátorů a sentimentu

Výstupem regresní analýzy je následující shrnutí. Statisticky významné regresní koeficienty jsou zvýrazněny červenou barvou (tabulka 8). Hladina významnosti $\alpha = 0,05$. Jak vyplývá z tabulky níže, v tomto modelu lze považovat za statisticky významné pouze proměnné RSI 9 a Stochastic. Všechny ostatní nezávisle proměnné mají poměrně vysoké p – hodnoty a nízké b^* , tedy nevysvětlují variabilitu závisle proměnné.

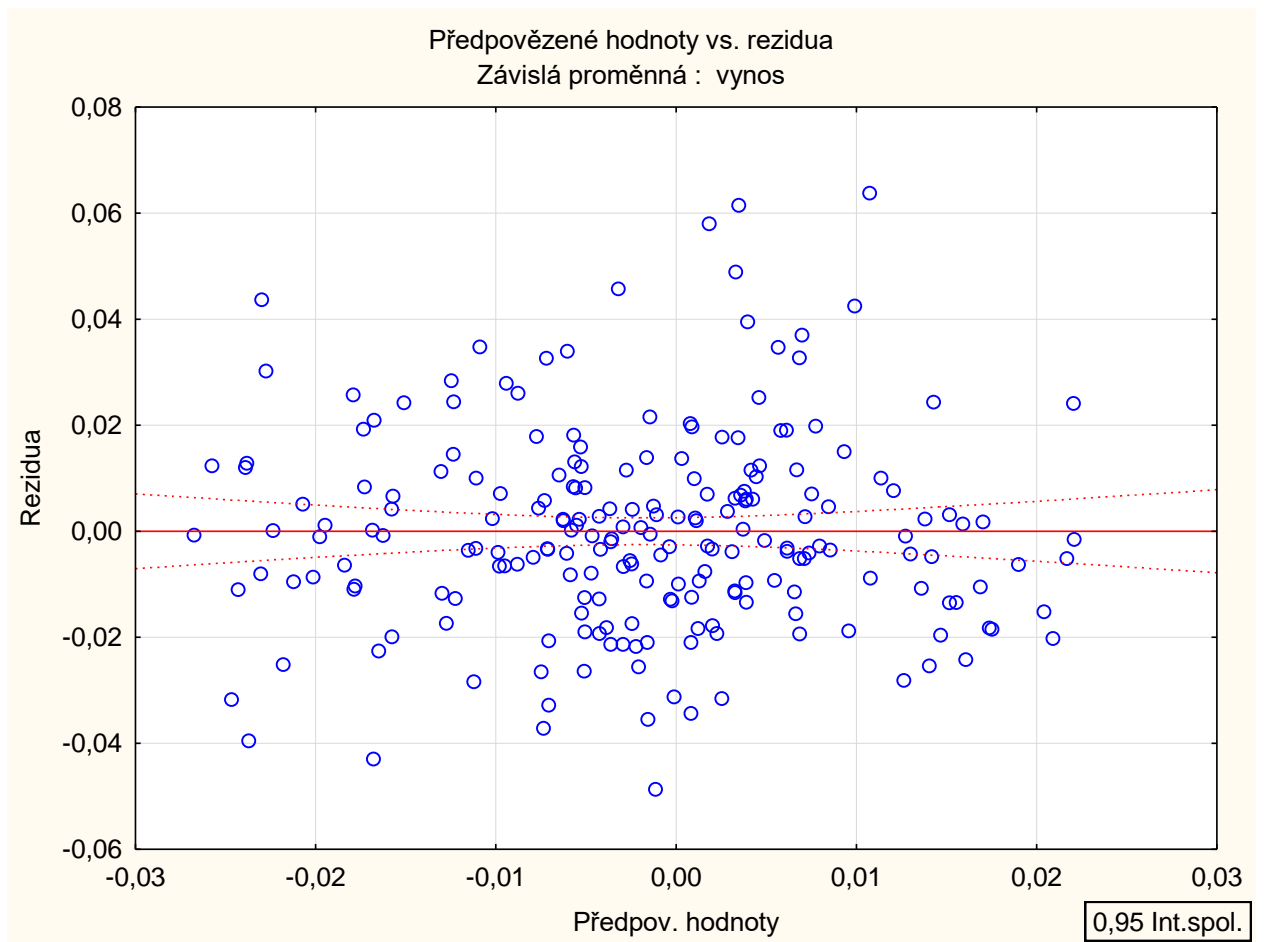
Tabulka 8: Výsledky regresní analýzy

N=216	Výsledky regrese se závislou proměnnou : vynos $R=,49894753$ $R^2=,24894864$ Upravené $R^2=,22367287$ $F(7,208)=9,8493$ $p<,00000$ Směrod. chyba odhadu : ,01913					
	b^*	Sm.chyba (z b^*)	b	Sm.chyba (z b)	t(208)	p-hodn.
Abs.člen			0,098535	0,070497	1,39771	0,163688
kod spol	-0,097250	0,069844	-0,000919	0,000660	-1,39239	0,165290
Bing Liu	0,019553	0,068066	0,004021	0,013999	0,28726	0,774199
Henry	0,032623	0,065833	0,007956	0,016056	0,49554	0,620742
Loughran, McDonald	-0,023278	0,069223	-0,004819	0,014330	-0,33627	0,737006
RSI 9	-0,515277	0,086732	-0,000596	0,000100	-5,94104	0,000000
Stochastic	0,678773	0,088924	0,000529	0,000069	7,63315	0,000000
Faktor MA	0,014028	0,068819	0,000306	0,001499	0,20383	0,838683

Zdroj: Vlastní zpracování

1. Největší vliv měl ukazatel Stochastic s hodnota koeficientu $b^* = 0,678773$. Rozdíl oproti ostatním ukazatelům je v tom, že pracuje s minimy a maximy ceny, a to v posledních čtrnácti dnech. Tyto údaje a časový horizont jsou tedy důležité.
2. Druhý vlivný ukazatel je Relative Strength Index devítidenní. Hodnota $b^* = -0,515277$. Záporný koeficient naznačuje nepřímou úměrnost, což je v souladu se smyslem ukazatele. Nízké hodnoty značí přeprodanost a tedy signalizují rostoucí trend a naopak vysoké hodnoty vypovídají o poklesu ceny. V porovnání s ostatními indikátory užitými v této práci je také spíše dlouhodobější.
3. Třetí největší význam měl identifikátor společnosti, Hodnota $b^* = -0,097250$, tato hodnota už je pod hranicí významnosti, nicméně z ní lze soudit, že situace nebyla u všech sledovaných společností identická.
4. Všechny proměnné zastupující sentiment se ukázaly jako nevýznamné. Stejně tak Faktor MA, který vysvětloval souhrnně klouzavé průměry. Tento fakt bude okomentován v závěru práce.

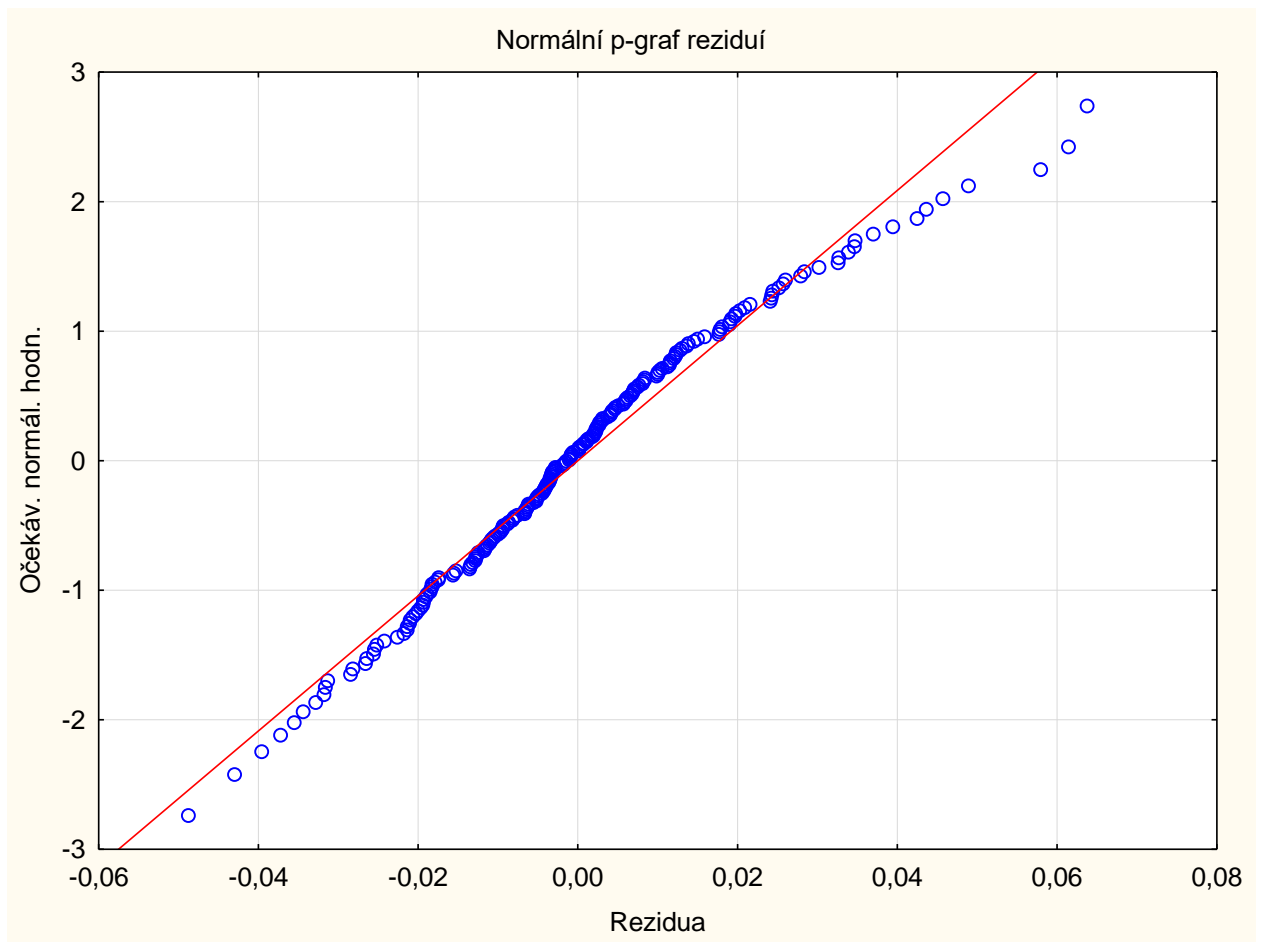
Regresní model má několik dalších předpokladů, které mají vliv na jeho přesnost. Pokud by se, v rámci regresního analýzy, mělo pokračovat v ověřování předpokladů modelu nebo si vyvolat další výstupy, stačí oživit si na liště záložku Reziduální analýza a pokračovat v práci [2]. Na závěr se tedy verifikuje použitý model i na základě jiných předpokladů. Nejprve se verifikuje model na základě reziduí (obrázek 16).



Obrázek 16: Předpovědi vs rezidua, software STATISTICA

Zdroj: Vlastní zpracování

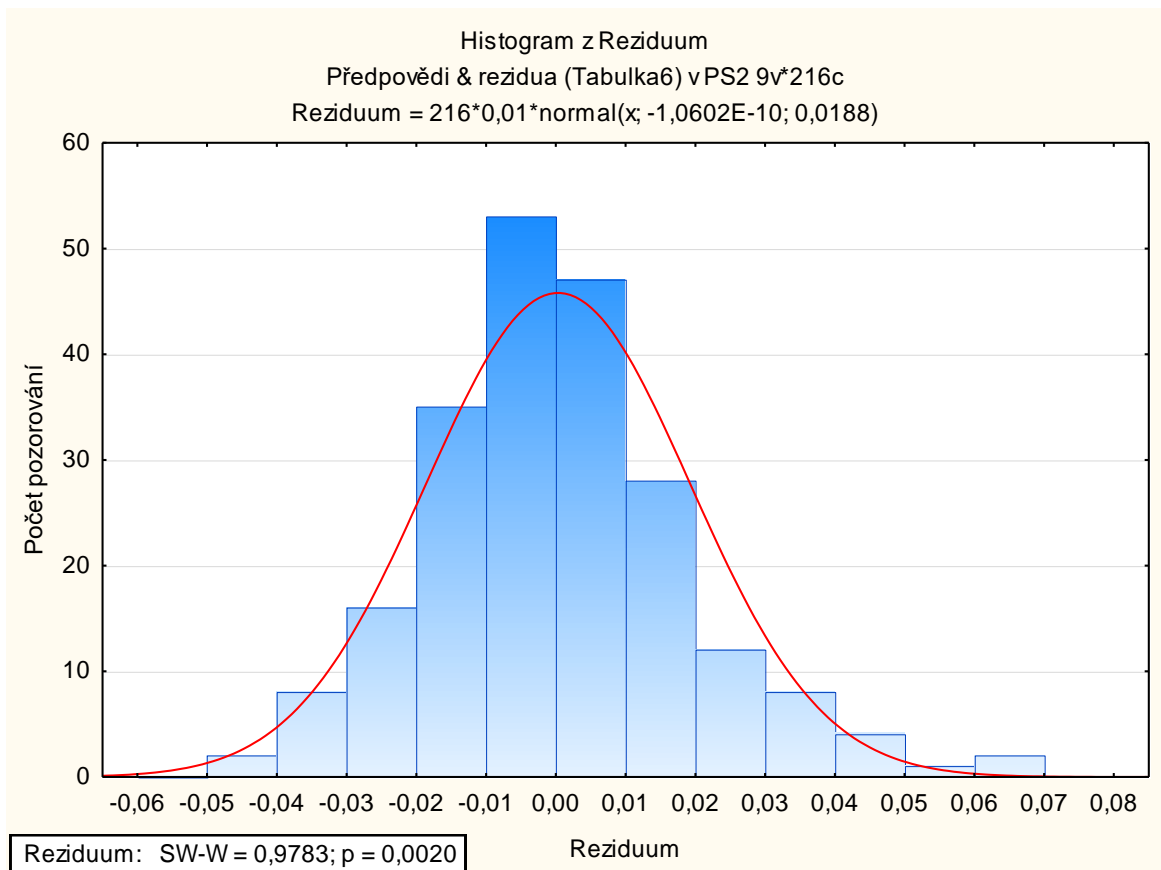
Na tomto grafu lze vizuálně ověřit následující předpoklady: Správně specifikovaný model – tedy jestli matematický tvar modelu je správně vybrán. Chybová složka má konstantní rozptyl (pro každé pozorování má příslušná složka vektoru E stejný rozptyl - tzv. podmínka homoskedasticity). A jednotlivé složky chybového vektoru jsou nekorelované [3].



Obrázek 17: Normální p – graf reziduí, software STATISTICA

Zdroj: Vlastní zpracování

Stejně tak můžeme na grafu (obrázek 17) pozorovat, zda rezidua pochází z normálního rozdělení. V krajních oblastech už se jednotlivé body od červené přímky nepatrně odchyľují, v ideálním případě by měly všechny body ležet na červené přímce nebo v její těsné blízkosti.



Obrázek 18: Histogram reziduí, software STATISTICA

Zdroj: Vlastní zpracování

Předpoklad normality je tedy vhodné ověřit ještě Shapiro – Wilkovým testem (obrázek 18). Výpočet reziduí byl aktivním vstupem pro histogram s Shapiro – Wilkovým testem, kde p hodnota vyšla menší než $\alpha=0,05$. I ze samotného grafu je patrné, že rozložení není zcela symetrické. Pro interpretaci výsledku to znamená, že odhady parametrů modelu a regresní rovnice jsou v pořádku, tam normalita není nepotřeba. Nicméně, předpoklad normality je důležitý pro určení významnosti regresních parametrů a pro správnost intervalů spolehlivosti [3]. S těmito údaji tedy nyní počítat nelze.

ZÁVĚR

Analýza sentimentu je nástroj užitečný ve velmi širokém spektru disciplín, finanční trhy jsou oborem, kde tato metoda v praxi ještě není příliš používána. Jde o náročné postupy, které dnes ještě zdaleka nejsou dovedeny k dokonalosti. V dnešní době rozhodně není problém s nedostatkem informací, naopak je třeba v přemíře informací najít vhodnou průzkumnou technikou ty správné. Analýza sentimentu založená na psaném textu se vždy bude muset vyvíjet alespoň tak rychle, jako jazyk sám. Vystihnout strojově myšlenku, kterou člověk zanesl do textu, je nesnadný úkol. V prostředí internetu se navíc objevuje řada dalších překážek a specifíků. Tato práce měla tematiku analýzy sentimentu v první řadě přiblížit, tento cíl byl splněn v první části. V návaznosti na tradiční metody analýz finančních trhů byly popsány přístupy autorů slovníků sentimentu pro účely analýz. Také byly shrnuty některé typické problémy, se kterými se analýza sentimentu potýká.

Druhým cílem bylo zjistit sentiment reálných komentářů v prostředí internetové diskuze a studovat jejich vývoj. Tento cíl byl splněn v další části, kde je popsán proces sběru a přípravy dat a následně i jejich zpracování. Pro analýzu bylo vybráno 8 akciových titulů. Finanční data i komentáře v příslušných diskuzích byly čerpány ze sociální sítě Yahoo Finance. Jednotlivé slova prošla analýzou, která na základě užití třech různých slovníků přiřadila výrazům pozitivní nebo negativní hodnotu. Jejich součty bylo ke každému dni ve sledovaném období přiřazeno skóre sentimentu. Tyto údaje byly zaneseny do grafu a spojeny do křivek, ze kterých byl patrný vývoj sentimentu v čase.

Posledním cílem bylo zjistit, zda jsou informace o sentimentu dostatečně spolehlivé pro predikci akciového vývoje. K jeho splnění byla využita regresní analýza. Ta byla doplněna z důvodu zvýšené korelace vstupů analýzou faktorovou. Proměnné založené na klouzavých průměrech byly sloučeny do jedné. Regresní analýza potom odhalila vliv sentimentu a technických indikátorů na vývoj ceny akcií. Ukázalo se, že významný vliv měl pouze oscilátor Stochastic a indikátor Relative Strength Index. Oblíbenost těchto ukazatelů tedy má zřejmě své opodstatnění. Vliv sentimentu nebyl natolik významný, aby umožnil predikci.

Je tedy třeba použití dokonalejších a náročnějších postupů, hlubší zkoumání mechanického zpracování jazyka a specifíků internetových diskuzí.

POUŽITÁ LITERATURA

- [1] BUDÍKOVÁ, M., STATISTIKA. Regresní analýza. [Online] Přírodovědecká fakulta, Masarykova univerzita, 2015. [Citace: 11. 04 2019.] Dostupné na: https://is.muni.cz/do/sci/UMS/el/popis_statistika/pages/regres-anal.html.
- [2] CR, STATSOFT. Možnosti vyhodnocení časových řad v softwaru STATISTICA. [Online] [Citace: 15. 04 2019.] Dostupné na: http://www.statsoft.cz/file1/PDF/newsletter/2013_04_02_StatSoft_Casove_rady_v_softwaru_STATISTICA.pdf.
- [3] CR, STATSOFT. Úvod do regresní analýzy. [Online] [Citace: 10. 04 2019.] Dostupné na: http://www.statsoft.cz/file1/PDF/newsletter/2014_26_03_StatSoft_Uvod_do_regresni_analyzy.pdf.
- [4] ČESKO 3.0. Creative Commons. Korelační a regresní analýza. [Online] 2013. [Citace: 12. 02 2019.] Dostupné na: https://wikisofia.cz/wiki/Korela%C4%8Dn%C3%AD_a_regresn%C3%AD_anal%C3%BDza. ISSN: 2336-5897.
- [5] FELDMAN, R. communications of the acm. Techniques and Applications for Sentiment Analysis. 2013, Sv. no. 4, doi:10.1145/2436256.2436274.
- [6] FIO BANKA. Technická analýza. Fio Banka. [Online] 2018. [Citace: 20. 02 2019.] Dostupné na: <https://www.fio.cz/akcie-investice/internetove-obchodovani/technicka-analyza>.
- [7] FXSTREET.CZ. FXstreet.cz. Forex slovník pojmů. [Online] 2009. [Citace: 11. 02 2019.] Dostupné na: <https://www.fxstreet.cz/forex-slovník-pojmu+market-sentiment-trzni-sentiment.html>.
- [8] GABROVŠEK P., ALEKSOVSKI D., MOZETIČ I., GRAČAR M. PLoS ONE. Twitter sentiment around the Earnings. Ljubljana : Department of Knowledge Technologies, 2017. 12. DOI:10.1371.
- [9] GOLDSTARWAY. Goldstarway. Exponenciální klouzavé průměry. [Online] 2016. [Citace: 04. 02 2019.] Dostupné na: <http://goldstarway.com/cz/prispevek/exponencialni-klouzave-prumery>.

- [10] HARMAN, R., KARLÍK, P., NEKULA, M., PLESKALOVÁ J. FAKTOROVÁ ANALÝZA. CzechEncy – Nový encyklopedický slovník češtiny. [Online] 2017. [Citace: 13. 04 2019.] Dostupné na: <https://www.czechency.org/slovník/FAKTOROV%C3%81%20ANAL%C3%9DZA>.
- [11] HARTMAN, O. Fxstreet. Technická analýza – oscilátory. [Online] 06. 12 2010. [Citace: 20. 02 2019.] Dostupné na: <https://www.fxstreet.cz/technicka-analyza-oscilatory.html>.
- [12] HAYES, A. Investopedia. Stochastic Oscillator Definition. [Online] 14. 04 2019. [Citace: 16. 04 2019.] Dostupné na: <https://www.investopedia.com/terms/s/stochasticoscillator.asp>.
- [13] HEBÁK, P., HUSTOPECKÝ, J., JAROŠOVÁ, E., PECÁKOVÁ, I. Vícerozměrné statistické metody. Praha: INFORMATORIUM, spol s. r. o., 2007. ISBN 978-80-7333-056-9.
- [14] HU, M., LIU B. Department of Computer Science. Mining and Summarizing Customer Reviews [Online] 2004. [Citace: 18. 02 2019.] Dostupné na: <https://www.cs.uic.edu/~liub/publications/kdd04-revSummary.pdf>.
- [15] HUŠEK, R. Ekonometrická Analýza. Praha: EKOPRESS, s. r. o., 1999. ISBN 80-86119-19-X.
- [16] JÍLEK, J. Akciové trhy a investování. Praha: GRADA Publishing, a.s., 2009. ISBN 978-80-247-2963-3.
- [17] JINDAL, N., LIU B. Department of Computer Science. Identifying Comparative Sentences in Text Documents. [Online] 2006. [Citace: 18. 02 2019.] Dostupné na: <https://www.cs.uic.edu/~liub/publications/sigir06-comp.pdf>.
- [18] JOVKOV, A. LYNX. Fundamentální analýza akcií: Praktický průvodce. [Online] 08. 02 2019. [Citace: 11. 02 2019.] Dostupné na: <https://www.lynxbroker.cz/vzdelavani/fundamentalni-analyza-prakticky/>.
- [19] KAMPS, J., MARX, M., MOKKEN, R. J., RIJKE, M. Language & Inference Technology Group. Using WordNet to Measure Semantic Orientations of Adjective. [Online] 2004. [Citace: 18. 02 2019.] Dostupné na: https://pure.uva.nl/ws/files/4089824/37038_file3940.pdf.
- [20] KOHOUT, P. Investiční strategie pro třetí tisíciletí. Praha: Grada Publishing, a. s., 2013. ISBN 978-80-247-5064-4.

- [21] KUBANOVÁ, J. Statistické metody pro ekonomickou a technickou praxi. Bratislava: Statis, 2004. ISBN 80-85659-37-9.
- [22] LI, J., HOVY, E. Reflections on Sentiment/Opinion Analysis. Determining the Sentiment of Opinions. [Online] 2004. [Citace: 18. 2 2019.] Dostupné na: http://delivery.acm.org/10.1145/1230000/1220555/p1367-kim.pdf?ip=89.111.70.134&id=1220555&acc=OPEN&key=4D4702B0C3E38B35%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35%2E6D218144511F3437&__acm__=1555439095_c2f2607fa74227166de6125b5c573ddc.
- [23] LIU, B. Sentiment Analysis: Mining Opinions, Sentiment, and Emotions. New York: Cambridge University Press, 2015. ISBN 978-1-107-01789-4.
- [24] LOUGHRAN, T., MCDONALD B. When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks. Journal of Finance, Forthcoming. 1, 04. 03 2010, ISSN 1540-6261.
- [25] MARR, B. Forbes. How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read. [Online] 21. 05 2018. [Citace: 11. 03 2019.] Dostupné na: <https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#2c5ac80360ba>.
- [26] MINER, G., ELDER, J., FAST, A., HILL, T., NISBET, R., DELEN, D. Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications. Waltham : Academic Press, 2012. ISBN 978-0123869791.
- [27] MUKHERJEE, A., LIU B. uic.edu. Aspect Extraction through Semi-Supervised Modeling. [Online] 2012. [Citace: 18. 2 2019.] Dostupné na: <https://www.cs.uic.edu/~liub/publications/ACL-2012-aspect-extraction.pdf>.
- [28] MURPHY C., LANGAGER CH. Investopedia. Dow Theory: The Three-Trend Market. [Online] 1999. [Citace: 06. 02 2019.] Dostupné na: <https://www.investopedia.com/university/dowtheory/dowtheory2.asp>.
- [29] ONLINEFINANCIALMARKETS.COM. OnlineFinancialMarkets.com. Elliott Wave Theory. [Online] 2008. [Citace: 06. 02 2019.] Dostupné na: <https://www.onlinefinancialmarkets.com/trading-commodities-online.html>.

- [30] PATRIA FINANCE, A.S. PATRIA.cz. Úvod do fundamentální analýzy. [Online] 1999. [Citace: 21. 02 2019.] Dostupné na: <https://www.patria.cz/akademie/analyzy-investice-fundamentalni-analyza.html>.
- [31] PLUMMER, T. Prognóza finančních trhů, psychologie úspěšného investování. Brno: Albatros Media a. s., 2014. ISBN 978-80-265-0063-6.
- [32] PUCHÝŘ, D. QUASTIC. Nejoblíbenější indikátor pod lupou: klouzavé průměry. [Online] 30. 09 2017. [Citace: 04. 02 2019.] Dostupné na: <https://quastic.cz/automaticke-obchodovani/nejoblibenejsi-indikator-pod-lupou-klouzave-prumery/>.
- [33] REJNUŠ, O. Finanční trhy. Praha: Grada Publishing a. s., 2014. 978-80-247-3671-6.
- [34] REJNUŠ, O. Finanční trhy: Učebnice s programem na generování cvičných testů. Praha: Grada Publishing a. s., 2016. ISBN 978-80-247-5871-8.
- [35] STATISTA. Most famous social network sites worldwide as of January 2019, ranked by number of active users (in millions). [Online] 11. 2 2019. <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>.
- [36] STATISTIKA A MY. Statistika a my. Nadějný růst ekonomiky ČR. [Online] 07. 08 2015. [Citace: 22. 02 2019.] Dostupné na: <http://www.statistikaamy.cz/2015/07/nadejny-rust-ekonomiky-cr/>.
- [37] ŠKALOUDOVÁ, A. Faktorová analýza. Stanovení počtu faktorů. [Online] Univerzita Karlova, 2010. [Citace: 11. 04 2019.] Dostupné na: http://kps.pedf.cuni.cz/skalouda/fa/stan_poc_faktoru.htm.
- [38] ŠKALOUDOVÁ, A., Faktorová analýza. Rotace faktorů. [Online] Univerzita Karlova, 2010. [Citace: 11. 04 2019.] Dostupné na: http://kps.pedf.cuni.cz/skalouda/fa/rotace_fak.htm.
- [39] VESELÁ, J. Investování na kapitálových trzích. Praha: Wolters Kluwer Česká republika, 2011. ISBN 978-80-7357-647-9.
- [40] XTB. xtb. Tržní sentiment a positioning. [Online] 2019. [Citace: 12. 02 2019.] Dostupné na: <https://www.xtb.com/cz/xtb-akademie/trzni-sentiment-a-positioning>.

SEZNAM PŘÍLOH

Příloha A: Vzorek neupravených dat titulu FB (vlastní zpracování)

Příloha B: Vzorek upravených dat titulu FB (vlastní zpracování)

Příloha A

Vzorek neupravených dat titulu FB

Net Flix

Net Flix19 days ago

Daniel Loeb's huge hedge fund dissolved his entire fakebook portfolio! :-o

Reply53

raggwead

raggwead19 days ago

Its down \$ 78 from highImagepafkufpafBy pafkufpaf on

Reply73

Omar

Omar19 days ago

and will hold for 2 years, that's about how much life this company has before it goes belly up

ReplyReplies (1)33

Omar

Omar19 days ago

I wont buy until \$98.55

Reply35

Deacon

Deacon19 days ago

FB almost back to a52 week low.

Reply1810

AntiShortie

AntiShortie19 days ago

I can't take the stupidity anymore. I have to stop entering these silly Yahoo chat rooms. Way too many morons in here.

ReplyReplies (2)84

Justin

Justin19 days ago

\$130 is inevitable.

ReplyReplies (1)812

Andy

Andy19 days ago

Oh Mary Joseph Abraham, have mercy on this lost sheep. For it does not know which direction to go. Please lead it to higher ground before it drowns itself in the bottomless abyss.

ReplyReplies (1)52

N.J.--MM-Trading Group -6

N.J.--MM-Trading Group -619 days ago

We have aFB price target lowered to 133

Reply59

Sean

Sean19 days ago

Legal spyware whose lifeblood is data. Except the blood is cancerous and will be subjected to more and more govt intrusion, regulation and control in the years to come.

ReplyReplies (2)23

peter

peter19 days ago

buy fb get rich

ReplyReplies (2)2412

peter

peter19 days ago

good entry price

ReplyReplies (1)2313

AntiShortie

AntiShortie19 days ago

Theres a 50/50 chance a stock will go up or down. Those of you who think you "called" this, are simply dumb. If there were 50/50 odds at winning the lottery, over 150 million people in the United States would be multi-millionaires.

Reply21

Zeke

Zeke19 days ago

WOW, dropping like a falling knife. Anyone buying FB in a falling knife scenario?

ReplyReplies (2)1310

Příloha B

Vzorek upravených dat titulu FB

12.11.2018; Daniel Loeb's huge hedge fund dissolved his entire fakebook portfolio! :-o

12.11.2018; Its down \$ 78 from highImagepafkufpafBy pafkufpaf on

12.11.2018; and will hold for 2 years, that's about how much life this company has before it goes belly up

12.11.2018; I wont buy until \$98.55

12.11.2018; FB almost back to a52 week low.

12.11.2018; I can't take the stupidity anymore. I have to stop entering these silly Yahoo chat rooms. Way too many morons in here.

12.11.2018; \$130 is inevitable.

12.11.2018; Oh Mary Joseph Abraham, have mercy on this lost sheep. For it does not know which direction to go. Please lead it to higher ground before it drowns itself in the bottomless abyss.

12.11.2018; We have aFB price target lowered to 133

12.11.2018; Legal spyware whose lifeblood is data. Except the blood is cancerous and will be subjected to more and more govt intrusion, regulation and control in the years to come.

12.11.2018; buy fb get rich

12.11.2018; good entry price

12.11.2018; Theres a 50/50 chance a stock will go up or down. Those of you who think you "called" this, are simply dumb. If there were 50/50 odds at winning the lottery, over 150 million people in the United States would be multi-millionaires.

12.11.2018; WOW, dropping like a falling knife. Anyone buying FB in a falling knife scenario?