

Univerzita Pardubice  
Fakulta ekonomicko-správní

Využití externích dat pro investiční modelování

Bc. Petr Šild

Diplomová práce

2019

## ZADÁNÍ DIPLOMOVÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Bc. Petr Šild**  
Osobní číslo: **E17536**  
Studijní program: **N6209 Systémové inženýrství a informatika**  
Studijní obor: **Pojistné inženýrství: Management finančních rizik**  
Název tématu: **Využití externích dat pro investiční modelování**  
Zadávací katedra: **Ústav matematiky a kvantitativních metod**

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

Cíl práce: Pomocí externích dat, jako jsou makroekonomická data, finanční data, či měnové agregáty, navrhnout investiční model. Daný model není ovlivněn subjektivními pocity investora a cílem tohoto modelu je určit podíl aktiv držných v portfoliu investora.

Osnova:

- Rešerše odborné literatury na téma modelování a investování.
- Analýza dostupných dat.
- Tvorba a implementování modelu.
- Analýza výkonnosti modelu.
- Vizualizace modelu pomocí vhodného softwaru.

Rozsah grafických prací: -  
Rozsah pracovní zprávy: cca 50 stran  
Forma zpracování diplomové práce: tištěná/elektronická  
Seznam odborné literatury:

FANTA, Jiří. Psychologie, algoritmy a umělá inteligence na kapitálových trzích. Praha: Grada, 2001. ISBN 80-247-0024-7.  
FOTR, Jiří a Jiří HNILICA. Aplikovaná analýza rizika ve finančním managementu a investičním rozhodování. 2., aktualiz. a rozš. vyd. Praha: Grada, 2014. ISBN 978-80-247-5104-7.  
MÁRTON, Peter a Norbert ADAMKO. Praktický úvod do modelovania a simulácie. Žilina: EDIS - vydavateľstvo ŽU, c2011. ISBN 978-80-554-0387-8.  
MEERSCHAERT, Mark M. Mathematical modeling. 4th ed. Waltham: Academic Press, 2013. ISBN 978-0-12-386912-8.  
MELOUN, Milan a Jiří MILITKÝ. Statistical data analysis: a practical guide : complete with 1250 exercises and answer key on CD. New Delhi: Woodhead Publishing India, 2011. ISBN 978-0-85709-109-3.  
ŠPAČEK, Miroslav. Pravděpodobnostní přístupy k analýze rizik investičních projektů a jejich využití v praxi. Plzeň: Nava, 2014. ISBN 978-80-7211-472-6.

Vedoucí diplomové práce:

  
RNDr. Jan Gogola, Ph.D.

Ústav matematiky a kvantitativních metod


Datum zadání diplomové práce: 3. září 2018

Termín odevzdání diplomové práce: 30. dubna 2019

  
doc. Ing. Romana Provaníková, Ph.D.

děkanka

L.S.

  
doc. RNDr. Bohdan Linda, CSc.  
vedoucí ústavu

V Pardubicích dne 3. září 2018

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.

V Pardubicích dne 30. 4. 2019

Bc. Petr Šild

Rád bych poděkoval vedoucímu práce RNDr. Jánu Gogolovi, Ph.D. za vstřícný přístup, pevné nervy, agilní přístup a cenné rady při zpracovávání diplomové práce.

## **ANOTACE**

V rámci této práce je sestaven model, který na základě tržních a externích dat identifikuje rizikové situace na trhu a dle nich doporučuje podíl aktiv v portfoliu. Sběr dat je provozován pomocí automatizovaných robotů a výsledky jsou poté vizualizovány v prostředí business intelligence nástroje, Qlik Sense.

## **KLÍČOVÁ SLOVA**

Investování, RPA, Python, Qlik Sense, model, datová analýza

## **TITLE**

Investing modeling by usage of external data

## **ANNOTATION**

In this final thesis, a model is created, which identifies risk situations on the market based on market and external data, and recommends a proportion of assets in the portfolio. Data collection is done by automated robots and the results are visualized in business intelligence tool, Qlik Sense.

## **KEYWORDS**

Investing, RPA, Python, Qlik Sense, model, data analytics

# OBSAH

SEZNAM OBRÁZKŮ.....	10
SEZNAM TABULEK .....	12
SEZNAM ZKRATEK .....	13
ÚVOD.....	14
1    INVESTOVÁNÍ NA KAPITÁLOVÝCH TRZÍCH.....	15
1.1    Fundamentální analýza.....	15
1.1.1    Globální fundamentální analýza.....	16
1.1.2    Odvětvová analýza .....	18
1.1.3    Firemní fundamentální analýza .....	21
1.2    Technická analýza.....	23
1.2.1    Dow Theory.....	24
1.2.2    Grafické metody .....	26
1.2.3    Analýza založená na technických indikátorech.....	30
1.3    Investiční omyly.....	42
1.3.1    Máte nadměrné sebevědomí a přehnaně důvěřujete ve své investiční schopnosti.....	43
1.3.2    Hledáte štěstí ve hvězdách.....	43
1.3.3    Myslíte si, že kvalita diverzifikace je určena počtem cenných papírů.....	44
1.3.4    Věříte v „horkou ruku“ manažerů fondu .....	44
2    DATA ANALYTICS & DATA SCIENCE.....	45
2.1    Co to je? .....	45
2.2    Životní cyklus.....	47
2.2.1    Klíčové role .....	47
2.2.2    Discovery.....	48
2.2.3    Data preparation .....	48

2.2.4	Model planning	49
2.2.5	Model building	49
2.2.6	Communicate results	49
2.2.7	Operationalize	49
2.3	Metody	50
2.3.1	Lineární regrese	50
2.3.2	Logistická regrese	52
2.3.3	Rozhodovací stromy	53
2.3.4	Simulace Monte Carlo	54
2.4	Python	56
2.4.1	Krátký tutoriál	56
2.4.2	Numpy	59
2.4.3	Pandas	60
3	ROBOTIC PROCESS AUTOMATION (RPA)	61
3.1	Úvod do automatizace	61
3.2	UI Path	62
4	PRAKTICKÁ ČÁST	64
4.1	Úvod do problému	64
4.2	Vlastní přístup	64
4.2.1	Sběr dat (Data collection)	66
4.2.2	Skladování dat (Data warehouse)	67
4.2.3	Příprava dat (Data preparation)	69
4.2.4	Návrh modelu	69
4.2.5	Příprava modelu	70
4.2.6	Vizualizace	79
4.3	Analýza výkonosti modelu	81



ZÁVĚR .....	83
POUŽITÁ LITERATURA .....	84

## SEZNAM OBRÁZKŮ

<b>Obrázek 1:</b> Graf vývoje životního cyklu odvětví .....	19
<b>Obrázek 2:</b> Vztah kurzu k vnitřní hodnotě v závislosti na efektivitě trhu .....	21
<b>Obrázek 3:</b> Jednotlivé trendové komponenty .....	25
<b>Obrázek 4:</b> Hranice podpory a odporu .....	26
<b>Obrázek 5:</b> Formace dvojitý vrchol a dvojitě dno .....	27
<b>Obrázek 6:</b> Formace trojitý vrchol a trojitě dno .....	27
<b>Obrázek 7:</b> Formace hlava a ramena .....	28
<b>Obrázek 8:</b> Formace diamant .....	28
<b>Obrázek 9:</b> Formace vlajka .....	29
<b>Obrázek 10:</b> Formace praporek .....	29
<b>Obrázek 11:</b> Vzestupné a sestupné klíny .....	30
<b>Obrázek 12:</b> Vývoj kurzu a 20 - denního vážného klouzavého průměru .....	32
<b>Obrázek 13:</b> Porovnání více klouzavých průměrů .....	33
<b>Obrázek 14:</b> Procentní pásma .....	34
<b>Obrázek 15:</b> Bollingerovy pásy .....	35
<b>Obrázek 16:</b> Vývoj Momentum pro Volkswagen .....	36
<b>Obrázek 17:</b> Vývoj Rate of change pro Volkswagen .....	37
<b>Obrázek 18:</b> Vývoj Relative strenght index pro Volkswagen .....	38
<b>Obrázek 19:</b> Vývoj MACD pro Volkswagen .....	39
<b>Obrázek 20:</b> Vývoj OBV pro Volkswagen .....	40
<b>Obrázek 21:</b> BI vs Data Science .....	46
<b>Obrázek 22:</b> Životní cyklus projektu .....	47
<b>Obrázek 23:</b> Grafická ilustrace lineární regrese .....	51

<b>Obrázek 24:</b> Logitová funkce .....	53
<b>Obrázek 25:</b> Výpočet $P_i$ v závislosti na počtu simulací .....	56
<b>Obrázek 26:</b> Kdy automatizovat? .....	62
<b>Obrázek 27:</b> UI Path prostředí .....	63
<b>Obrázek 28:</b> Vlastní přístup.....	65
<b>Obrázek 29:</b> Tržní data v UI Path.....	66
<b>Obrázek 30:</b> Dataset v SQL.....	68
<b>Obrázek 31:</b> Klasifikační report IT.....	75
<b>Obrázek 32:</b> Klasifikační strom IT .....	76
<b>Obrázek 33:</b> Qlik Sense: závislost konstanty a hranice na výnosu .....	78
<b>Obrázek 34:</b> Vizualizace: akciový model.....	79
<b>Obrázek 35:</b> Vizualizace: akciový model – zdravotní péče .....	80
<b>Obrázek 36:</b> Vizualizace: sektorový model.....	80
<b>Obrázek 37:</b> Vizualizace: celý trh .....	81
<b>Obrázek 38:</b> Porovnání kumulovaného výnosu od roku 2004 .....	82
<b>Obrázek 39:</b> Porovnání volatility od roku 2004 .....	82

## **SEZNAM TABULEK**

<b>Tabulka 1:</b> Hospodářské expanze a recese v USA do roku 2001 .....	16
<b>Tabulka 2:</b> Výčet předbíhajících indikátorů .....	18
<b>Tabulka 3:</b> Porovnání rozhodovacích stromů.....	54

## **SEZNAM ZKRATEK**

S&P	STANDARD AND POOR'S
CEO	CHIEF EXECUTIVE OFFICER
SQL	SIMPLE QUERY LANGUAGE

# ÚVOD

V rámci diplomové práce, bude čtenář seznámen s třemi poslední dobou velmi často skloňovanými tématy a jednou stálíci. Poté bude seznámen, jak by tato témata šla propojit a prakticky využít.

Prvním často skloňovaným tématem je datová analýza. Přímo nebo nepřímo se s ním setkal každý z nás, jelikož společnosti mají přístup k velkým objemům dat a hledají způsob, jak je efektivně využít. Proto také týmy datových analytiků patří mezi ty nejrychleji rostoucí. Dalším tématem je robotizace či automatizace. Opět velmi časté téma, jelikož není dostatek pracovní síly. V době krize se situace moc nezmění, jelikož firmy budou chtít ušetřit finanční prostředky, s čímž se také automatizace pojí. Posledním z často skloňovaných témat je vizualizace. Díky vizualizaci je výrazně snazší pochopit data a je možné nalézt vzory, které dříve nebylo možné spatřit.

Jako poslední téma nutno zmínit investování a kapitálové trhy, které se lidstva drží již desítky let. Na tomto místě by bylo vhodné vzpomenout výrok Isaaca Newtona, poté co ztratil jmění v Jihomořské bublině: „Mohu spočítat hybnost vesmírných těles, ale šílenství lidí nikoliv.“

Práce je rozdělena do čtyř kapitol. První je věnována investování na kapitálových trzích a rozebírá metody technické analýzy. V druhé kapitole je čtenář seznámen se základními principy datové analýzy, ve čtvrté zase s principy automatizace. V rámci páté kapitoly je sestaven model, který analyzuje rizika na kapitálových trzích a dává doporučení investorovi skladby portfolia.

# 1 INVESTOVÁNÍ NA KAPITÁLOVÝCH TRZÍCH

Kapitálové trhy již po delší dobu lákají pozornost investorů, regulátorů i akademiků. Investoři pod vidinou zajímavých zisků jsou ochotni se vzdát současné hodnoty svých finančních prostředků, za účelem vyšší budoucí hodnoty, která je však nejistá. Právě z důvodu zpřesnění té nejisté budoucí hodnoty vznikla řada analytických postupů, díky kterým se dá alespoň trochu odlišit investování od hraní rulety či blackjacku. Čtenář bude v rámci této kapitoly seznámen s některými metodami fundamentální a technické analýzy a závěru bude popsáno několik investičních omylů, jichž se investoři běžně dopouštějí. [25]

## 1.1 Fundamentální analýza

Pokud by někdo hledal velice komplexní a složitý přístup, jak zanalyzovat akcii, je tu právě pro něj fundamentální analýza. Fundamentální analýza má vskutku značný záběr, nezkoumá pouze čistě firemní faktory, jako zadluženost, historické zisky, dividendy, rentabilitu či likviditu. Vedle těchto faktorů zkoumá i globální faktory, které působí na trh jako celek či odvětvové faktory, působící na určité odvětví, v kterém se firma pohybuje. Fundamentální analýza tedy zkoumá akcie na třech úrovních:

- Globální fundamentální analýza
- Odvětvová analýza
- Firemní fundamentální analýza

Při analyzování akcie, na těchto třech úrovních, se analytik snaží přijít na jednu klíčovou otázku: „Je akcie podhodnocena či nadhodnocena? Nebo eventuálně ohodnocena na správné úrovni? Vedle této otázky je dokonce schopen odpovědět: „Proč je akcie takto tržně oceněna a co z toho plyne pro budoucí vývoj?“ [25] [22] [16] [3]

### 1.1.1 Globální fundamentální analýza

Tabulka 1: Hospodářské expanze a recese v USA do roku 2001

Dno	Vrchol expanze	Doba trvání expanze (v měsících)	Doba trvání recese (v měsících)	Propad HDP
XI-27	VIII-29	21	12	-
III-33	V-37	50	43	-
VI-38	II-45	81	13	-
X-45	XI-48	37	8	-
X-49	VII-53	45	11	3,6%
V-54	VIII-57	39	10	1,9%
IV-58	IV-60	24	8	3,2%
II-61	XII-69	107	10	0,5%
XI-70	XI-73	36	11	0,1%
III-75	I-80	58	16	3,4%
VII-80	VII-81	12	6	2,2%
XI-82	VIII-90	94	16	2,8%
III-91	II-00	108	7	1,3%

*Zdroj: upraveno podle [25]*

Koukněme se na tabulku 1, kde jsou hospodářské recese a expanze v USA do roku 2001. Mít tuto tabulku v roce 1927, tak by se kdokoliv mohl stát „nechutně“ bohatým. Vzhledem k tomu, že věšteckou kouli ani stroj času zatím nikdo nevynalezl, musíme tedy trávit čas globální fundamentální analýzou a čelit nástrahám zákeřných černých labutí (naprosto nečekaná událost s velkým negativním dopadem).

Hlavním posláním globální fundamentální analýzy je analyzovat vliv celé ekonomiky a trhu na hodnotu analyzované akcie. K analýze využívá makroekonomické agregáty, jako jsou úrokové míry, HDP, inflace či peněžní zásoba. Z pravidla dává smysl vždy analyzovat dvě veličiny a jejich vzájemný vliv jedné na druhou. Podrobným zkoumáním již byly objeveny určité vazby. Kupříkladu jedna z nejznámějších je vliv úrokové míry na vývoj akciových kurzů. [25] [22]

Zde dle studií, zmiňovaných Bernsteinem [1], platí negativní vztah s korelačním koeficientem -0,85. Tedy pokud rostou úrokové míry, má to negativní dopad na akciové kurzy a ty pak klesají. Tato vazba je poměrně logická, pokud rostou úrokové sazby, dluhopisy (či jiné investiční instrumenty, jejichž výnos se odvíjí od úrokových sazeb) se stanou více žádané a vzroste po nich poptávka, což má za následek zvýšení nabídky akcií (jelikož investoři musí čerpat finanční prostředky), což „in ceteris paribus“ vede k poklesu cen akcií. [25]



Alternativním vysvětlením by mohlo být, že investoři oceňují akcie pomocí diskontovaných budoucích příjmů. Tedy pokud vzroste úroková míra, kterou diskontují, pak klesne vnitřní hodnota akcie, což má za následek pokles poptávky po akciích, dokud se tržní ocenění nepřiblíží jejich vnitřní hodnotě stanovené pomocí diskontního modelu. [25] [22] [16]

Naopak pozitivní vztah byl identifikován mezi reálným výstupem z ekonomiky pohybem akciových kurzů. Tento vztah má avšak jednu malou vadu na kráse. Bylo zjištěno, že v rámci krátkého či střednědobého horizontu, akciové kurzy předbíhají vývoj reálného výstupu z ekonomiky (ať již měřený pomocí HDP či indexu průmyslové produkce) o tři až devět měsíců. Tudíž pokud by se někdo snažil prognózovat vývoj akciových kurzů pomocí vývoje HDP, jeho prognóza by byla o tři až devět měsíců zpožděná, jinými slovy nepoužitelná. [25] [22] [16]

Dalším pozitivním vztahem, který byl identifikován, je vztah mezi peněžní nabídkou a akciovými kurzy. Bohužel dle novějších studií je identifikováno, že tento vztah postupně slábne. Nicméně, jak by se dal tento vztah vysvětlit? Při zvýšení peněžní nabídky investoři budou mít pocit nadbytku likvidity a tuto nadbytečnou likviditu „uklidí“ mimo jiné i do akcií, čímž se zvýší jejich poptávka, a tedy i rovnovážná cena. Peněžní nabídka je indikátorem předbíhající, tudíž k predikci akciových kurzů je možné ho použít. [25] [22] [16]

Obecně vzato dává smysl používat pouze předbíhající indikátory. Jejich celkový výčet je možné najít v následující tabulce:

**Tabulka 2:** Výčet předbíhajících indikátorů

Pořadí	Předbíhající (vedoucí indikátory)
1.	Peněžní nabídka (M2)
2.	Index spotřebitelských očekávání
3.	Změna v cenách materiálů
4.	Nová stavební povolení
5.	Objednávky strojního zařízení a vybavení (změna)
6.	Objednávky zásob (změna v zásobách na skladě a v objednávkách)
7.	Průměrný počet hodin, odpracovaných týdně dělníky
8.	Průměrné týdenní požadavky na dávky pro nezaměstnané
9.	Změna v počtu poskytnutých podnikových a spotřebních úvěrů
10.	Změna v objemu likvidních aktiv
11.	Ziskovost (zisková marže)
12.	Peníze a finanční toky (cash flow)
13.	Mezní změna v zaměstnanosti

*Zdroj: upraveno podle [25]*

### 1.1.2 Odvětvová analýza

V rámci odvětvové analýzy se již analyzují pouze vlivy spojené s konkrétním odvětvím na kurz akcie. Konkrétně tak jde o životní cyklus odvětví, citlivost na hospodářský cyklus či regulatorní vliv. [25] [22]

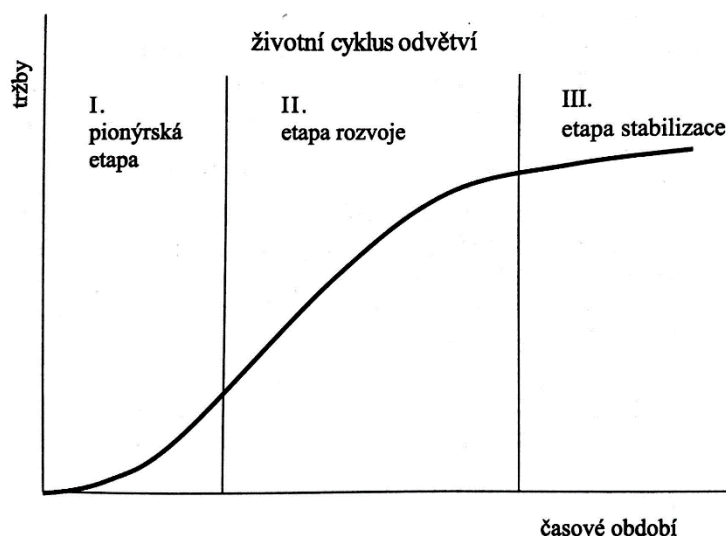
#### 1.1.2.1 ŽIVOTNÍ CYKLUS ODVĚTVÍ

První, zahajovací fázi, je **pionýrská fáze**. Je typická prudkým růstem poptávky po výrobcích firem, jelikož nesou něco nového, inovativního, čímž je také spojen velký růst kurzů akcií těchto firem. Velká poptávka po výrobcích má však také za následek velký příval konkurence, které lákají zisky což ústí v určitou volatilitu, jak některé firmy příval konkurence neunesou. [25]

Další fázi je **fáze rozvoje**, kdy se odvětví stabilizuje a firmy, přeživší pionýrskou fází si budují postavení na trhu a expandují. V této fázi je možné se setkat s již nižší volatilitou, nicméně klesá míra růstu tržeb firem a cen produkce vlivem zvýšené konkurence. [25]

Finální **fáze stabilizace** již je charakteristická stabilními zisky i kurzy, jelikož už tu jsou významní hráči na trhu. Problémem však jsou ceny produkce, které klesají, což má za následek, že některé firmy odvětví opouštějí. Odvětví se pak může vyvíjet dle dvou scénářů: buď pokračujícím poklesem cen a postupným úpadkem odvětví, nebo převratnou inovací, která

oživí a „zrestartuje“ odvětví, čímž způsobí opětový nárůst zisků a poslání celého odvětví opět do pionýrské fáze. [25]



**Obrázek 1:** Graf vývoje životního cyklu odvětví

*Zdroj: [25]*

#### 1.1.2.2 CITLIVOST ODVĚTVÍ NA HOSPODÁŘSKÝ CYKLUS

Vývoj tržeb, zisků a kurzů nereaguje v rámci všech odvětví stejně na hospodářský cyklus. Právě podle intenzity reakce jsou identifikovány následující odvětví:

U **Cyklických odvětví** je možné pozorovat, že vývoj jejich tržeb, zisků či kurzů relativně kopíruje vývoj hospodářského cyklu. Nejvýkonnější jsou v období konjunktury, a naopak největšího úpadku zažívají během krize. Logicky z toho můžeme určit, že jejich beta koeficient se pohybuje v intervalu  $<1;\infty$ ). Logicky do této skupiny patří například luxusní statky, které v době krize až tak moc lidí nekupuje (pozn.: existují i výjimky, jako například Luis Vuitton, jehož tržby v období velké hospodářské krize v letech 2008-2009 nějak výrazně neklesly). Další odvětví, které je možné do této skupiny zařadit jsou odvětví statků, jejichž spotřebu lze odložit, tedy například stavebnictví, hotelnictví, automobilový průmysl, strojírenství a další. [25] [22] [16]

V případě **Neutrálních odvětví** již není možné pozorovat nějakou výraznější vazbu zisků na hospodářský cyklus. Typická je produkce nezbytných statků, jako například potravinářský průmysl. Lidé musí jíst, ať již je konjunktura nebo krize a pokud se daří ekonomice, tak se těžko rozhodnou, že najednou budou jíst dvakrát tolik. Dalším příkladem je například farmaceutický průmysl, nebo základní drogistické zboží. [25] [16]

**Anticyklická odvětví** jsou typická tím, že největších zisků dosahují v době recese. Tento jev je poměrně neobvyklý, jelikož v období recese se celkově spotřeba snižuje. Tyto odvětví však prosperují, jelikož nabízejí zastupitelný levnější substitut. Avšak toto platí pouze po omezenou dobu, dokud na trh nepříjde substitut jiný, lepší. Příkladem můžou být například videopůjčovny v 90. letech, které představovaly levnější substitut chození do kina. V nynější době internetu, by se však videopůjčovnám nedařilo jak době recese, tak ani v době konjunktury. [25] [16]

V rámci otázky citlivosti odvětví na trh je třeba také vzít v úvahu poměr variabilních a fixních nákladů. Firmy s větším objemem variabilních nákladů jsou schopny lépe reagovat na útlum v tržbách snížením celkových nákladů. Pokud ovšem vezeme v úvahu takovou energetiku, tak například ČEZ nemůže přestat udržovat polovinu infrastruktury a propustit polovinu techniků či jaderných fyziků a nechat Temelín svému osudu. [25]

Dalším faktorem je celková zadluženost společnosti. Pokud firma používá velké množství cizích zdrojů ke svému financování, tak sice v období konjunktury do určité míry, tím zvyšuje svůj ukazatel ROE (Return on equity), avšak v období krize tím může být dosti nestabilní a dostat se do finančních problémů. [25]

### 1.1.2.3 VLIV REGULATORNÍCH ORGÁNŮ

Zde je možné identifikovat 2 základní druhy regulací. Jedna s pozitivním dopadem na schopnost analyzovat a predikovat budoucí kurzy s čímž častokrát bývá spojen i nárůst zisků a kurzů akcií, a druhá se stejným dopadem, ale negativním.

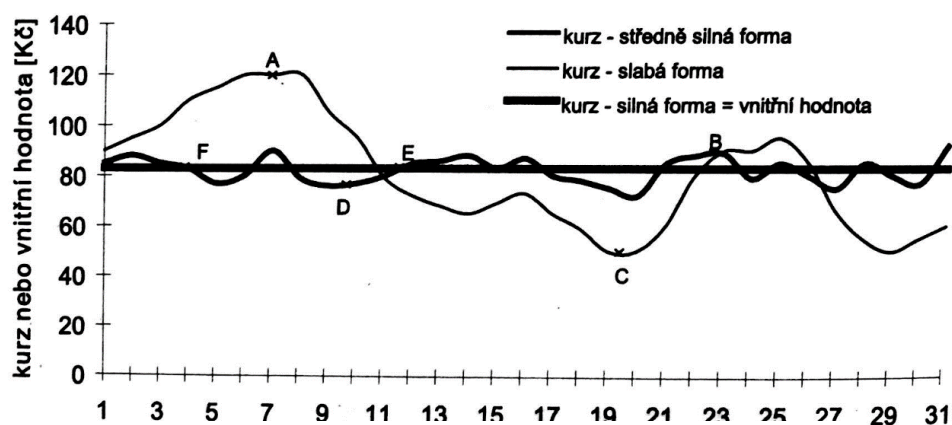
Mezi ty s pozitivním dopadem patří kupříkladu **omezování vstupu do odvětví**, což vede k oligopolní struktuře, která má z pravidla vyšší zisky a je stabilnější. Příkladem můžou být třeba banky či pojišťovny. Další pro investory pozitivní regulací je **regulace cenotvorby** kdy stát „zastropuje“ cenu výrobků, kde na tom nějaký sociální, alokační či redistribuční motiv. Tento efekt pakt z pravidla vede opět k vyšší stabilitě firem, včetně jejich akciových kurzů. Příkladem může být třeba distribuce energie či vody. Výrazně pozitivní vliv má také **preferenční odvětví ze strany vlády**. Firmám poté narostou tržby ještě o vládní zakázky, čímž vzrostou zisky i akciové kurzy. [25] [22]

Pak tu jsou naopak negativní vlivy regulací, jako třeba **poskytování dotací a subvencí**. Sice se může na první pohled zdát, že jde o pozitivní vliv, ale v konečném důsledku častokrát tyto dotace nejsou efektivně vyžity v businessu a kýžený pozitivní efekt se nedostaví. Je tedy

velice obtížné tento efekt predikovat. Zde je typickým příkladem zemědělství (např. Agrofert). Další negativní regulací může být **regulatorní opatření vyvolávající dodatečné náklady firem** či **stanovení základních pravidel pro hospodářskou soutěž**. V obou případech jde o dodatečné zvyšování nákladů firem, ať už formou daní, kolků či sankcí, které nelze moc predikovat dopředu. [25] [22]

### 1.1.3 Firemní fundamentální analýza

Poslední úroveň fundamentální analýzy je firemní fundamentální analýza. Principem firemní fundamentální analýzy je pomocí firemních faktorů určit vnitřní hodnotu akcie a tu pak porovnat vůči tržní ceně a na základě toho určit, jestli se jedná o akcii nadhodnocenou, či podhodnocenou. Firemní fundamentální analýza pracuje s myšlenkou, že vlivem nedokonalé efektivnosti trhu, diskrepancí mezi oceněním profesionálními investory (zde je pro ně používán termín „smart money“) a drobnými investory („laikové“) a dalších psychologických či technických faktorů, dochází ke kolísání tržního ocenění akcie vůči její vnitřní hodnotě, což je možné spatřit na obrázku 2. [25] [22]



**Obrázek 2:** Vztah kurzu k vnitřní hodnotě v závislosti na efektivitě trhu

*Zdroj: [25]*

Na obrázku 2 je zobrazeno ocenění akcií v závislosti na efektivnosti trhu, kde při silně efektivním trhu se tržní ocenění takřka rovná stabilní vnitřní hodnotě akcie, pak na málo efektivním trhu kurz akcie značně kolísá kolem její vnitřní hodnoty. [25]

Je zde možné identifikovat body A a B, kdy je akcie nadhodnocená a není výhodné takovouto akcii držet. Naopak je možné vidět body C a D, kdy je akcie podhodnocená a tuto akcii je vhodné nakoupit. V případě bodů E a F je akcie správně tržně oceněna a takovou akcii je vhodné držet. [25]

Další otázkou je, jak správně stanovit vnitřní hodnotu akcie? Pro tento účel byla vyvinuta řada metod.

Jednou skupinou jsou metody, které pracují s diskontováním budoucích příjmů plynoucí s držby akcie. Mezi ně řadíme **dividendový diskontní model**, jehož principem je diskontování všech budoucích dividend. Buď je brána v úvahu nekonečná doba držby, čímž z modelu odpadá hodnota akcie za  $n$  let a pracuje se zde s mírou růstu dividend. Další možností je model konečné doba držby, do kterého vstupují dividendy, které budou vyplaceny za následujících  $n$  let a hodnota akcie za  $n$  let. Dále do této skupiny řadíme **ziskové modely**. Tyto modely jsou založeny na principu, že se vezme historický vztah ceny a nějakého ziskového ukazatele (EBIT, EBT, EAT, tržby...) a tento vztah se promítne do budoucnosti. Tedy na základě předpokládaného ziskového ukazatele se odhadne předpokládaná cena akcie. Zde je nejpoužívanější ukazatel P/E ratio – poměr ceny a čistého zisku. **Cash flow modely** oproti předchozím modelům nesledují pouze tok peněz od akcionářů a k nim, ale promítají zde i toky peněz v rámci společnosti. Výhodou je, že lze takto ocenit akcie společností, které vyplácejí malé nebo dokonce žádné dividendy. Konstrukci takového cash flow modelu popisuje rovnice 1. [25] [22] [16]

**Rovnice 1:** Konstrukce Free Cash flow to Equity

$$FCFE_0 = EAT(\text{čistý zisk}) + \text{odpisy} - \text{investiční výdaje} - \text{změna pracovního kapitálu} \\ - \text{splátky dluhu} + \text{nové emise dluhových instrumentů}$$

*Zdroj: Upraveno dle [25]*

Druhou skupinou modelů jsou bilanční modely, které se oproti těm předchozím nedívají do Výkazu zisku a ztrát („výsledovky“), ale do rozvahy. Na základě toho stanoví hodnotu firmy, posléze jedné akcie, a tu pak porovnávají s tržní cenou. Jedním způsobem je například **Účetní hodnota** (Book value). Ta je na konstrukci nejjednodušší, jelikož představuje rozdíl mezi aktivy a cizím kapitálem. Dalším způsobem je **substanční hodnota**, která vychází z účetní hodnoty, nicméně jednotlivé položky aktiv jsou přeceněny tržními hodnotami. Nakonec je vhodné zmínit **likvidační hodnotu**. Tu je možné získat, pokud by firma ukončila svoji činnost a veškerá aktiva by se zpeněžila dle aktuálních cen, odečetly by se náklady na prodej a splatily by se veškerý cizí kapitál. [25] [16]

## 1.2 Technická analýza

Jednou z metod, jak porozumět, co se na trzích děje a ideálně se tak trochu snažit predikovat, co by se dít mohlo, je technická analýza. Můžeme ji definovat jako grafickou techniku, která zachycuje časové řady cen a objemu obchodních transakcí. [25] [3]

Její kořeny sahají až do 18. století na asijské rýžové trhy, kde její metody hojně využíval úspěšný obchodník, Munehisa Honma (jeden z prvních průkopníků). Poté na něj navázal Charles H. Dow a spolu se svými následovníky dali vzniku ucelené Dow theory, která je považovaná za východisko technické analýzy. [25]

V souvislosti s technickou analýzou určitě stojí za zmínku úspěšný obchodník s akciemi a komoditami W. D. Gann, jehož principy byly shrnuty do Gannovy teorie, která lze využít k predikci, kdy lze očekávat změnu trendu. Dalším, kdo se snažil rozpoznat změnu trendu byl R.N. Elliot, který navázal na Dowovu teorii a stojí za vznikem teorie Elliotových vln, v polovině 19. století. [25]

Jaká je tedy podstata technické analýzy? Jde o prognózu pohybů kurzů a vývoj trendů, a to na základě studia grafů. Tyto grafy pak představují základní a hlavně nezbytný nástroj technické analýzy. Dle názorů mnoha /někoho/ zabývajících se technickou analýzou „pouze ti, kteří studují grafy, jsou schopni rozeznat, co přijde“. Jelikož přirozeně na trhu vedle fundamentálních faktorů působí také vlivy psychologické a iracionální, které nelze adekvátně předpovídat, techničtí analytici v průběhu času shrnuli základní principy a předpoklady, které využívají. A to konkrétně do tří tezí. [25] [11]

### 1. Vývoj na trhu diskontuje všechno

Veškeré nové a relevantní informace jsou promítnuty do ceny akciového titulu. Změna avšak není okamžitá, ale je postupná. Zakládá tedy vzniku trendů, které se snaží techničtí analytici analyzovat. Zajímá je tedy pohyb sám o sobě a nikoliv příčina toho pohybu. [25]

### 2. Existují vzory v pohybu kurzů

Techničtí analytici věří v existenci vzorů v pohybu kurzů a jejich cílem je včasné identifikování pomocí různých metod. Právě včasná a správná identifikace vzoru je klíčová k dalšímu prognózování vývoje kurzu. [25]

### 3. Historie se opakuje

Nemalou řádku let (více než jedno století) investovali techničtí analytici do pozorování, kategorizace a porovnání vzorů. Dospěli tedy k závěru, že tyto vzory jsou v čase neměnné, jelikož i podstata lidské psychiky je v čase neměnná. [25] [11]

Jak je všem dozajista jasné, všechny podmínky této teorie nejsou splněny v plném rozsahu. Například: „historie se opakuje“. Na tuto tezi by bylo vhodné odpovědět ano i ne. Zcela jistě je možné pozorovat určité cykly, které se opakují, avšak problémem je, že ne vždy úplně přesně. Výsledkem je určitá nepřesnost a zkreslenost výsledků technické analýzy. [25]

Jako další omezení je zde samotný princip technické analýzy, kdy je cílem odpovědět na otázku: „kdy se co stane“. Je tedy vhodné ji použít na načasování obchodů, avšak nikoliv pro výběr titulů akcií do portfolia. Pro tento problém je vhodné použít buď již dříve zmíněné metody fundamentální analýzy, či Markowitzovu teorii portfolia. [25]

Techničtí analytici vyvinuli spoustu metod. Tyto metody můžeme rozřadit do 2 skupin, na metody grafické a technické indikátory. Avšak tyto metody dávají smysl použít pouze na efektivním trhu. Na méně efektivních trzích je jakákoliv analytická činnost vycházející ze všeobecně dostupných informací odsouzena k neúspěchu. Příkladem může být například Newyorská burza před rokem 1929, kde probíhalo velké množství nekalých praktik a manipulací s trhem a do toho davové nadšení a masové investování na úvěry. Výsledkem byl pád newyorské burzy 24.10.1929, který zaskočil i technické analytiky. [25] [11] [16]

#### 1.2.1 Dow Theory

Touto teorií položil Charles Henry Dow, její autor, základy moderní technické analýzy. C.H.Dow a index DJIA – Dow Jones Industrial Average není pouhá shoda jmen, nýbrž Dow stojí i za zrodem tohoto indexu, jehož předchůdcem byl Railroad Average – index jedenácti vybraných akcií (devět byly železniční společnosti). Základem této teorie je následujících 7 principů:

*Indexy v sobě odrážejí všechny relevantní informace*

*Pohyb lze rozdělit do 3 trendových komponentů:*

a) Primární trend

Dlouhodobý trend za období jednoho a více roků



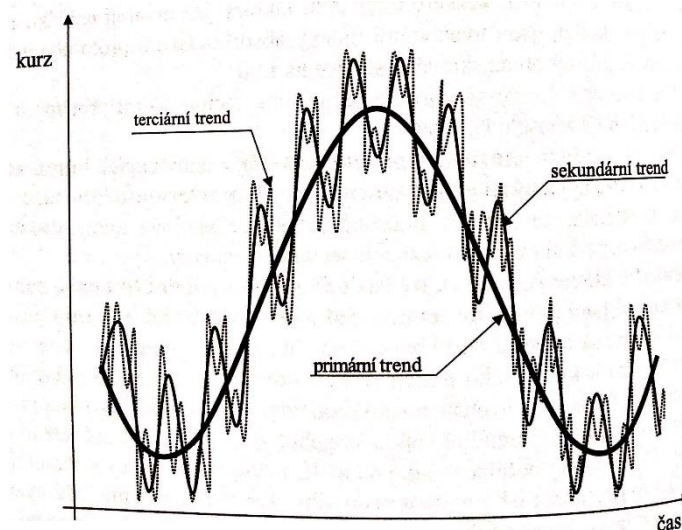
b) Sekundární trend

Méně významná kolísání za období 3 měsíců až jednoho roku.

c) Terciární trend

Krátkodobé fluktuace kurzu akcie za časové období v rámci několika málo dnů.

Tyto trendové komponenty jsou lépe vyobrazeny na Obrázku 3



**Obrázek 3:** Jednotlivé trendové komponenty

Zdroj: [25]

*Budoucí vývoj trendu lze odvodit z minulé tržní situace*

V rámci tohoto principu Dow rozlišuje 2 trendy: rostoucí trend býčí, tedy kdy každý další vrchol je vyšší než ten minulý, a trend medvědí, kdy naopak každý další dno je nižší, než dno minulé.

*Primární trendy obsahují tři fáze*

- Akumulační - kdy nakupují nejlépe informovaní investoři poté, co trh dosáhl svého dna
- Rostoucí – zde nakupují zejména institucionální investoři a trh masivně roste
- Distribuce – v této fázi nakupuje široká veřejnost (Aleš Michl tuto fázi u Bitcoinu popsal: „Už o něm mluví i moje babička a to má starou Nokii“). Po vyčerpání trh přechází do medvědího trendu.

*Akciové indexy se musí navzájem potvrzovat*

*Objemy obchodů musí potvrzovat trend*

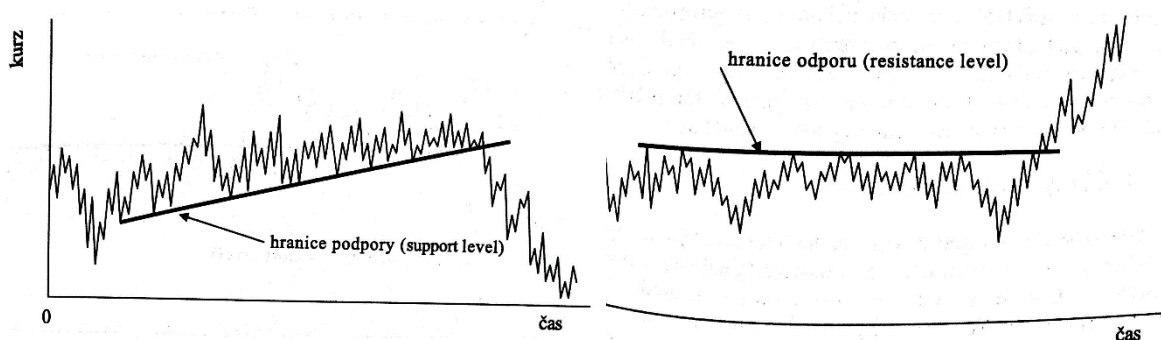
*Nastoupený trend trvá až do doby, kdy nastoupí jasné signály, že došlo ke změně*

[25] [11] [16]

### 1.2.2 Grafické metody

Hlavní podstatou grafických metod je odhalení standardizovaných formací a na základě nich stanovení buď kupních či prodejních signálů v případě analýzy jednotlivých titulů akcií, nebo identifikace primárního trendu v případě analyzování celého trhu. Formace se dále dělí na formace potvrzující trend (např. trojúhelníková formace či vlajková formace) a formace vedoucí ke změně trendu (M-W formace či V formace). Speciálním druhem grafických metod je Elliotova teorie vln, která vychází z přírodních zákonů a Fibbonacchiho posloupnosti. [25] [11] [16]

Obvykle se pracuje s dvěma hranicemi: horní hranice odporu a spodní hranice podpory. Pokud se kurz pohybuje blízko hranice odporu, investoři přestávají nakupovat akcie, jelikož kurz vzrostl příliš vysoko a akcie se staly příliš drahými. Tento trh bývá označován jako překoupený. Za následek to má pokles poptávky po titulu a následný pokles kurzu. Hranice podpory funguje více versa. Akcie se staly příliš levnými, což má za následek růst poptávky, jelikož investoři rádi „nakupují ve výprodejích“. Důsledkem je na přeprodaném trhu růst kurzu. Speciálním případem je, když dojde k „proražení hranice“. V případě horní hranice se akcie budou obchodovat za vyšší cenu a obě hranice se posunou směrem nahoru. Pro dolní hranici toto opět platí více versa. Hranice jsou pro lepší srozumitelnost vyobrazeny na obrázku 4. [25] [11] [16]



**Obrázek 4:** Hranice podpory a odporu

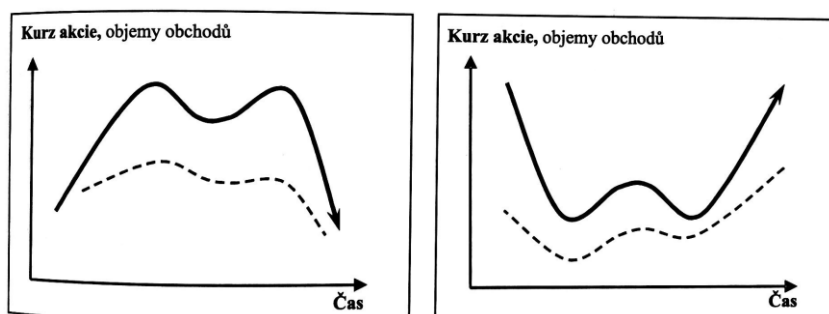
*Zdroj: [25]*

### 1.2.2.1 REVERZNÍ (ZVRATOVÉ) FORMACE

Tyto formace, jak již napovídá název, slouží k identifikaci změn v trendu. Právě těmto formacím věnují techničtí analytici největší pozornost, jelikož představují buď příležitost, nebo hrozbu. Níže je tedy popsáno několik vybraných z nich. [16]

#### *Formace dvojitý vrchol a dvojitě dno*

Dvojitý vrchol a dvojitě dno patří mezi nejdůležitější formace, jelikož se objevují poměrně zřídka a signály jsou značně spolehlivé. Jak je možné vidět na obrázku 5, typické jsou shodnou výškou svých vrcholů a objemem obchodů, který téměř kopíruje tvar formace, až na druhý vrchol, respektive dno, které musí být výrazně nižší, respektive vyšší než předchozí. [16]

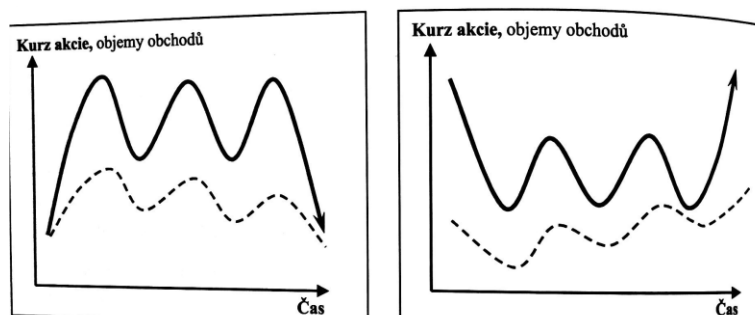


**Obrázek 5:** Formace dvojitý vrchol a dvojitě dno

*Zdroj: [16]*

#### *Formace trojitý vrchol a trojitě dno*

Tyto formace jsou poměrně vzácné a velice podobné formacím dvojitý vrchol a dvojitě dno a formacím hlava a ramena (této formaci je věnován další odstavec). Rozdílem od formace dvojitý vrchol a dno je jeden extra vrchol či dno navíc. Od formace hlava a ramena se liší shodnou výškou všech vrcholů/den. [16]

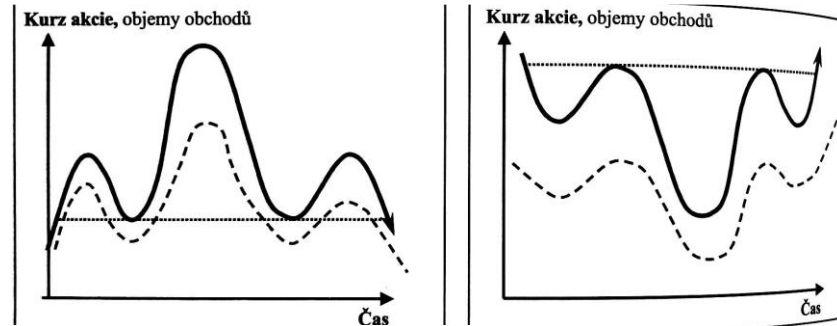


**Obrázek 6:** Formace trojitý vrchol a trojitě dno

*Zdroj: [16]*

### *Formace hlava a ramena*

Jako nejspolehlivější formace je považována právě tato. Jedná se o kombinaci formací vrchol a dno, avšak s proměnlivou velikostí. Dále je potřeba sledovat objem obchodů, pokud v případě reverzní formace s medvědího trhu na býčí dojde k překročení linie krku při nízkém objemu obchodů, dost možná se jedná o planý signál. [16]

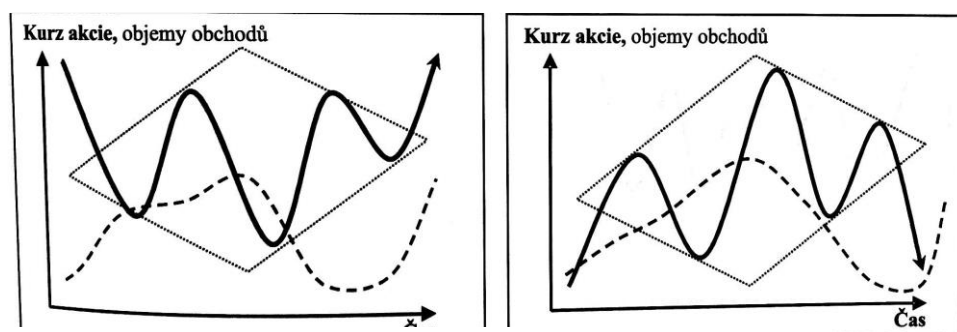


**Obrázek 7:** Formace hlava a ramena

*Zdroj: [16]*

### *Formace diamant*

Formace diamant se také nevyskytuje zrovna často. Obvykle se objevuje jen při velmi vysokých kurzech, jsou s ním tedy spojeny i vysoké objemy obchodů. V případě této formace pod pojmem diamant je potřeba si představit nepravidelný rovnoběžník složený ze dvou symetrických trojúhelníků. Tento rovnoběžník pak pochopitelně není tvarem grafu, nýbrž tvarem ve kterém je vývoj kurzu akcie vepsaný. [16]



**Obrázek 8:** Formace diamant

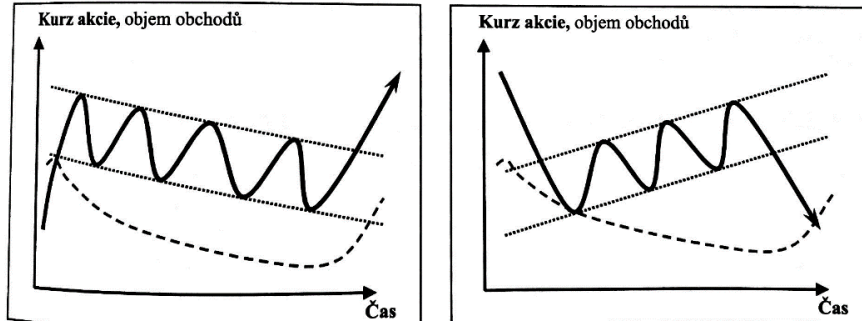
*Zdroj: [16]*

#### **1.2.2.2 KONSOLIDAČNÍ FORMACE**

Konsolidační formace na rozdíl od těch zvrátových mají za cíl potvrdit trend. Smyslem je tedy určit, že stávající sestupný, či rostoucí trend bude pokračovat i do budoucna. Těchto formací existuje velké množství, nicméně dále je uvedeno několik z nich. [16]

### *Formace vlajka*

Formace představuje krátkodobou korekci v minulém prudkém růstu či poklesu. Nicméně po opuštění této formace dojde k původnímu poklesu či růstu. [16]

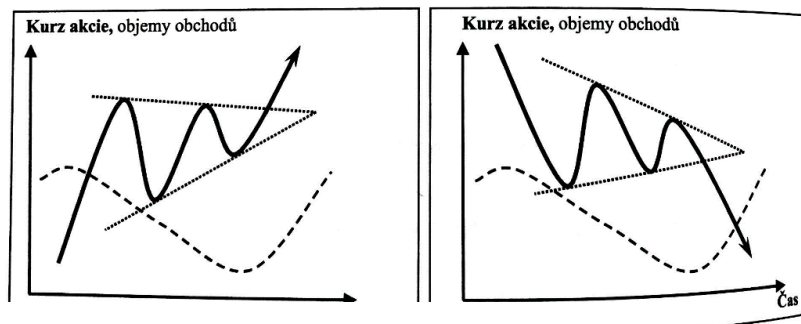


**Obrázek 9:** Formace vlajka

*Zdroj: [16]*

### *Formace praporek*

Praporek stejně jako vlajka představuje korekci v rostoucím, či klesajícím trendu. Jediným rozdílem je, že se hraniční přímky, v rámci kterých kurz osciluje, postupně sbíhají. [16]

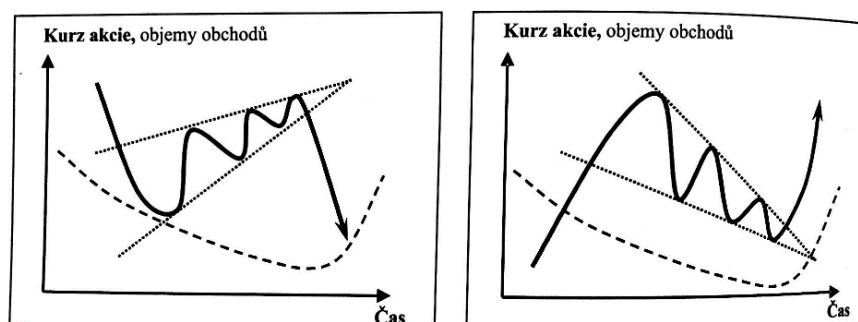


**Obrázek 10:** Formace praporek

*Zdroj: [16]*

## *Vzestupné a sestupné klíny*

Tato formace je opět poměrně podobná těm předešlým. Rozdílem zde avšak je, že kurz v případě klesajícího trendu zamíří výrazně nahoru a v případě rostoucího trendu výrazně dolů. Při tom také dochází ke snižování objemu obchodů. K prolomení klínu dochází při opětovém růstu obchodů. [16]



**Obrázek 11:** Vzestupné a sestupné klíny

*Zdroj: [16]*

### **1.2.3 Analýza založená na technických indikátorech**

V jakékoliv literatuře se můžeme dočíst o technických indikátorech něco jako: „metody, které analyzují tržní, objemové nebo cenové charakteristiky“. Reálně si je můžeme představit tak, že v grafických metodách jsme měli grafy, kde jsme hledali obrazce a podle nich předpovídali, co by se tak mohlo dít na trhu. Nyní již budeme aplikovat trochu matematických metod, jako jsou například klouzavé průměry a ty pak budeme dále analyzovat. Tyto metody můžeme dále dělit na dvě velké skupiny metod. Jedna sloužící k analýze jednotlivých akciových titulů a druhá sloužící k analýze trhu jako celku. [25] [16]

#### **1.2.3.1 ANALÝZA JEDNOTLIVÝCH AKCIOVÝCH TITULŮ**

Hlavními skupinami technických indikátorů jsou klouzavé průměry, pásmová analýza, oscilátory a cenové či objemové indikátory.

#### ***Klouzavé průměry***

Klouzavé průměry jsou jedním z nejpoužívanějších technických indikátorů. Jejich hlavní výhodou je, že umí vyhlazovat krátkodobé výkyvy kurzu, naopak jejich nedostatkem je jejich zpoždění při vytváření signálů. Obecně platí, čím dlouhodobější klouzavý průměr je použit, tím bude indikovat signály s větší přesností, ale zároveň s větším zpožděním. Pro krátkodobější klouzavé průměry toto tvrzení platí vice versa. [25] [16] [3]

Při použití klouzavého průměru je klíčové zvolit jeho správnou délku. Při analýze dlouhodobých trendů (tedy identifikaci býčího či medvědího trhu) je vhodné použít dlouhé klouzavé průměry, například 200-denní, aby nedávaly předčasné signály o změně trendu. Při analýze střednědobých trendů je zase vhodnější zvolit krátkodobější, například 50-denní a pro krátkodobé trendy použít několikadenní klouzavé průměry. [25] [16] [3]

Můžeme uvést několik druhů klouzavých průměrů:

#### *Jednoduchý klouzavý průměr*

Jednoduchý klouzavý průměr je poměrně často využíván pro svoji jednoduchost. Vznikne pomocí jednoduchého aritmetického průměru 10 po sobě jdoucích hodnot časové řady. V praxi avšak občas může být nevýhodná jeho zásadní vlastnost, a ta, že všem datům přiřazuje stejnou váhu. [25] [16]

#### *Vážený klouzavý průměr*

Pokud k hodnotám jednoduchého klouzavého průměru přiřadíme ještě vektor vah, který nebude pouze konstanta, získáme tím vážený klouzavý průměr. Z pravidla dává smysl přiřadit tím menší váhy, čím se více vzdalujeme do minulosti, tedy například vezmeme desetidenní klouzavý průměr a hodnotě 10 dní staré dáme váhu 1, hodnotě staré 9 dní váhu 2 a tak dále až hodnotě z minulého dne přiřadíme váhu 10. Následně musíme všechny váhy vydělit jejím součtem (55), abychom dostali sumu finálních vah rovnou jedné. Alternativním přístupem může být přiřazení nejvyšší váhy prostření hodnotě a tím nižším, čím by se více blížily do současnosti či minulosti. [25] [16]



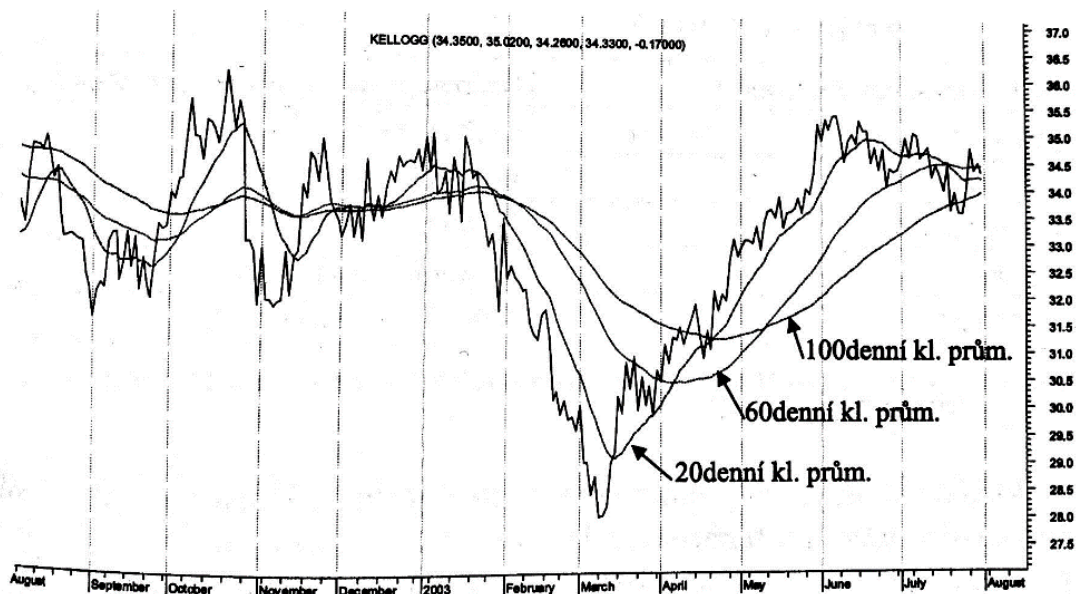
**Obrázek 12:** Vývoj kurzu a 20 - denního váženého klouzavého průměru

*Zdroj: [25]*

Po sestrojení váženého klouzavého průměru je třeba pozorovat vztahy s denní křivkou. Jako ilustrační příklad nám zde může posloužit Obrázek 12, kde vidíme vývoj akciového kurzu, oproti jeho 20 - dennímu klouzavému průměru u akcie Kellogg. Nutno dodat, že na svislé ose je cena v USD a na vodorovné je znázorněn čas. Dojde-li k protnutí váženého průměru křivkou zespoda, znamená to nákupní signál a naopak. Jako doplněk je třeba sledovat objemy obchodů. Pokud k protnutí křivky dojde při malém objemu obchodů, z pravidla jde o planý signál a nelze tomu přikládat velký význam. [25] [16]

Dalším přístupem je porovnání dvou a více klouzavých průměrů vůči sobě a generování nákupních či prodejních signálů vzájemným protnutími. Příkladem tohoto přístupu je například metoda MACD (Moving average convergence divergence).





**Obrázek 13:** Porovnání více klouzavých průměrů

*Zdroj: [25]*

#### *Exponenciální klouzavý průměr*

Exponenciální klouzavé průměry vychází z vážených klouzavých průměrů. Rozdílem však jsou váhy, které u vážených klouzavých průměrů směrem do minulosti klesají lineárně, u exponenciálních klouzavých průměrů klesají exponenciálně. [25] [16]

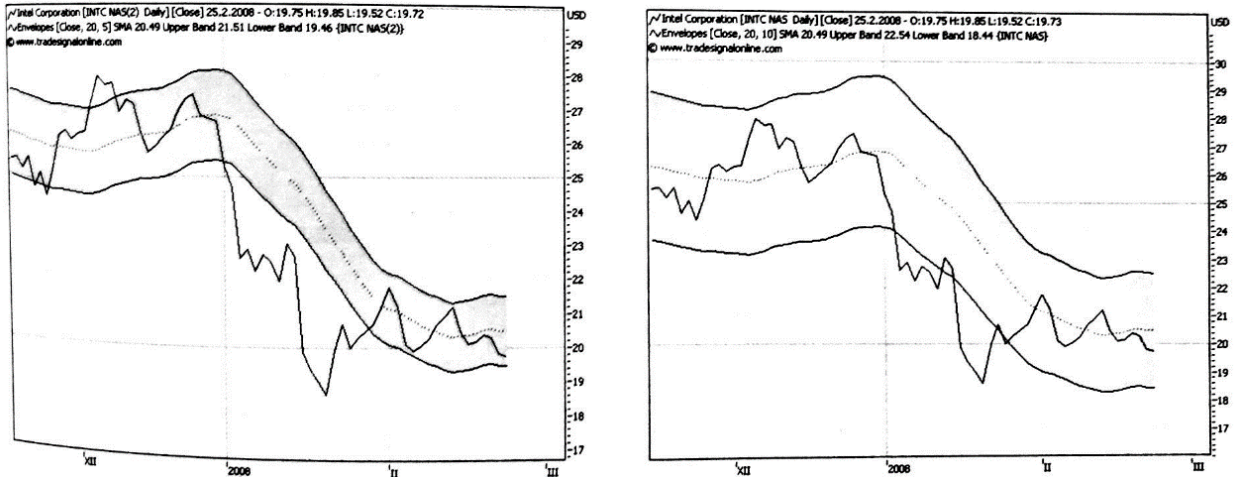
#### *Pásmová analýza*

Pásmová analýza představuje praktické využití klouzavých průměrů. Základem jsou tři křivky, a to klouzavý průměr a takzvané obálky, které tuto křivku obklopují. Horní hranice, představují hranici odporu, a spodní hranice, představující hranici podpory, jsou od klouzavého průměru stejně vzdáleny, a to buď podle procentní konstanty, nebo dle volatility dané akcie. [25] [16]

#### *Procentní pásma*

Jak již bylo řečeno, procentní pásma se konstruují pomocí klouzavého průměru, který je středem tohoto pásma a poté hranicí podpory a odporu, které jsou vzdálené určitý počet procent (v ilustrovaném příkladu to je 5 % a 10 %). Pro tyto pásma se v praxi používá výraz obálky (envelop). V případě, že křivka vývoje kurzu prorazí horní hranici odporu a poté se navrátí zpět, jedná se o prodejní signál. Identifikace nákupního signálu probíhá vice versa. [25] [16]

Pro pásma je klíčové správně nastavit jejich šíři. Pokud budou moc široká, tak nám nákupní či prodejní signály nepřinesou žádné a když už přinesou, tak příliš pozdě. Naopak pokud budou příliš úzká, tak nákupních či prodejních signálů přinesou mnoho a řada z nich nebudou správné. [25] [16]

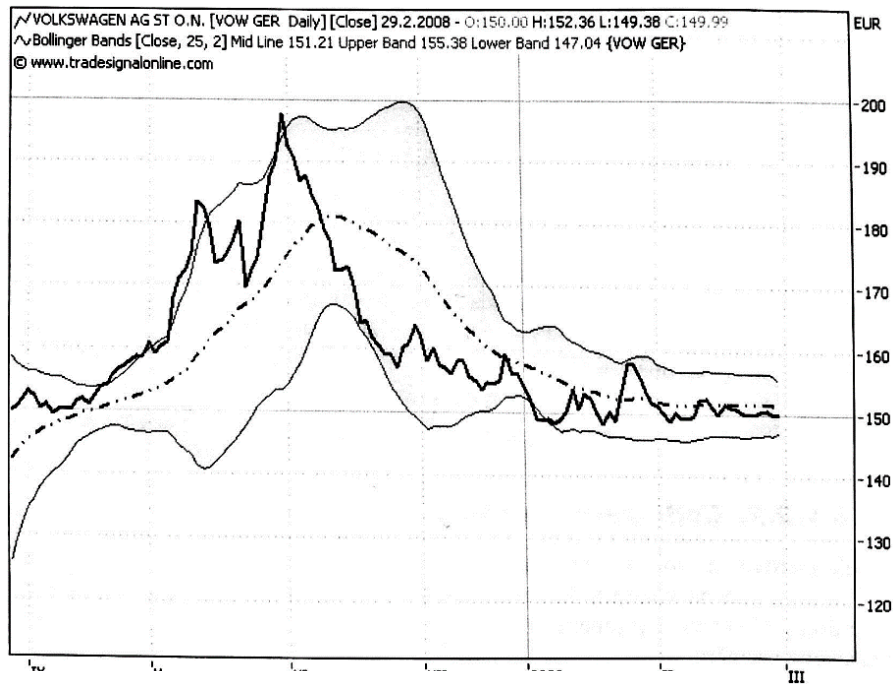


**Obrázek 14:** Procentní pásma

*Zdroj: [16]*

### *Bollingerovy pásy*

Analytik John Bollinger vylepšil pásmovou analýzu o to, že již není tak závislá na tom, jestli širší pásma je dobře nastavena. Šířka pásma neurčil jako konstantu, která se odvíjí od nejlepšího odhadu analytika, nýbrž na výši směrodatné odchylky. Bollingerem je doporučeno použít 20-denní klouzavý průměr a pásmo o šířce  $\pm 2$  směrodatné odchylky. Výsledkem je, že širší pásma je proměnlivá podle kolísání kurzu akcie, tedy širší pásma lépe reaguje na aktuální situaci na trhu. Pokud je kurz velmi rozkolísaný, zvětší se také širší pásma a nedochází ke generování velkého množství planých signálů. Naopak při malém kolísání kurzu dojde k zúžení pásma, a tato fáze se nazývá jako konsolidační (potvrzení trendu).



**Obrázek 15:** Bollingerovy pásy

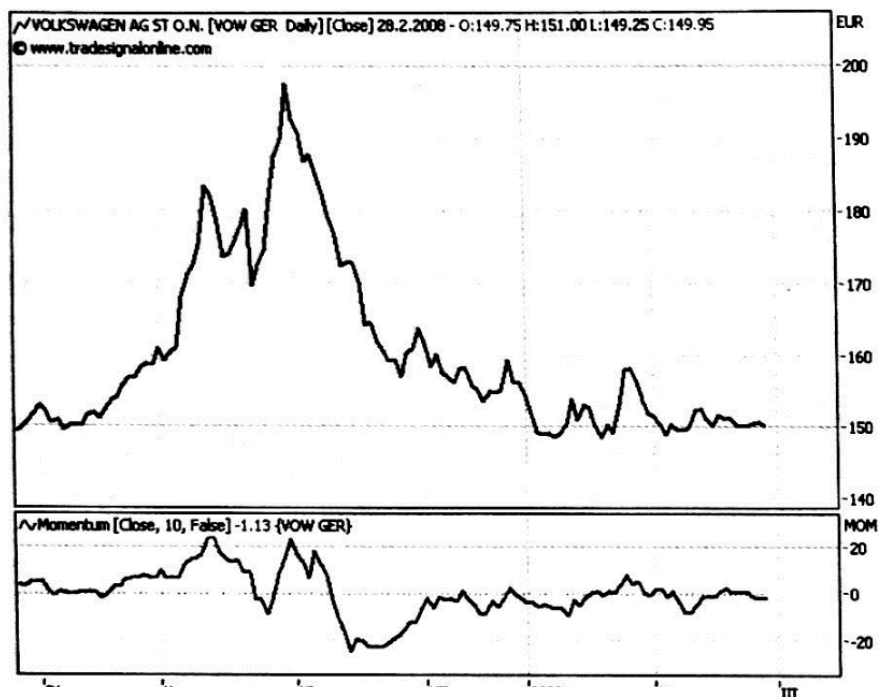
*Zdroj: [16]*

### *Oscilátory*

Oscilovat znamená kolísat kolem něčeho. Tedy i oscilátory kolísají, a to buď kolem vymezené úrovně, nebo v rámci stanoveného pásma. Smyslem oscilátorů je měřit sílu nastoupeného trendu, intenzitu jeho kolísání a pravděpodobnost jeho změny. Jejich konstrukce je opět poměrně jednoduchá, jen je potřeba zvolit správné hranice, které nám posléze udávají signály k nákupu či prodeji. Ač je oscilátory možné použít k analýze medvědího či býčího trhu, doporučuje se je spíše užít k analýze postranního trhu (vedlejších trendů), kde naopak ostatní metody zpravidla selhávají.

### *Momentum*

Momentum je nejjednodušším oscilátorem, jelikož k jeho konstrukci je potřeba udělat pouze podíl nebo rozdíl 2 hodnot s určitým časovým rozestupem. Smyslem je zobrazit zpomalení nebo zrychlení nastoupeného trendu. Při rozdílu hodnot vzniká absolutní momentum, při podílu momentum relativní, přičemž doporučené časové zpoždění se pohybuje mezi 10-12 dny. Logicky pak z výsledného grafu vyplývá, že pokud sílí Momentum, pak trend nabývá na síle, a naopak pokud slábne, slábne i trend a je možné očekávat v budoucnu jeho korekci. [25] [16]



**Obrázek 16:** Vývoj Momentum pro Volkswagen

*Zdroj: [16]*

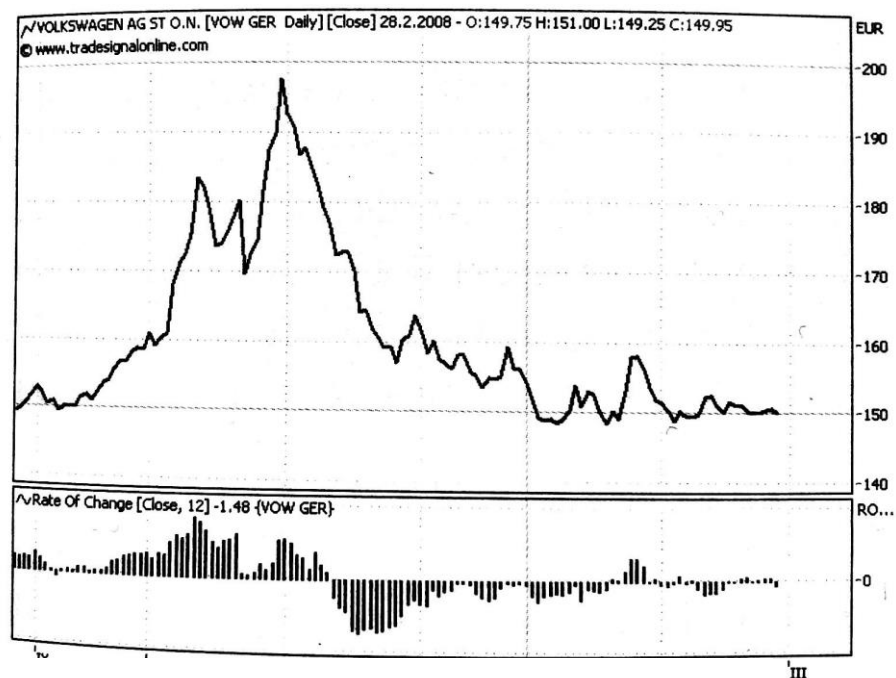
*Rate of Change (ROC)*

**Rovnice 2: Rate of change**

$$ROC_T(n) = \frac{Kurz_t - Kurz_{t-n}}{Kurz_{t-n}} \cdot 100$$

*Zdroj [16]*

Pokud zkombinujeme dohromady absolutní a relativní momentum dostaneme indikátor ROC. Ilustrativně tedy vezmeme rozdíl mezi kurzy s určitým časovým zpožděním a ten rozdíl ještě dále vydělíme kurzem s časovým zpožděním. Nákupní impulzy poté indikujeme, pokud index klesá a nabývá velmi nízkých hodnot a naopak. [25] [16] [3]



**Obrázek 17:** Vývoj Rate of change pro Volkswagen

*Zdroj: [16]*

### *Relative strenght index (RSI)*

Index relativní síly je poměrně jednoduchý na konstrukci, avšak poněkud složitější na interpretaci. Zkonstruovat index je možné dle vzorce 3, kde  $U(n)$  je počet kladných kurzových změn za období  $n$  a  $D(n)$  je počet záporných kurzových změn za období  $n$ , v určitém časovém okamžiku  $t$ . [25] [16] [3]

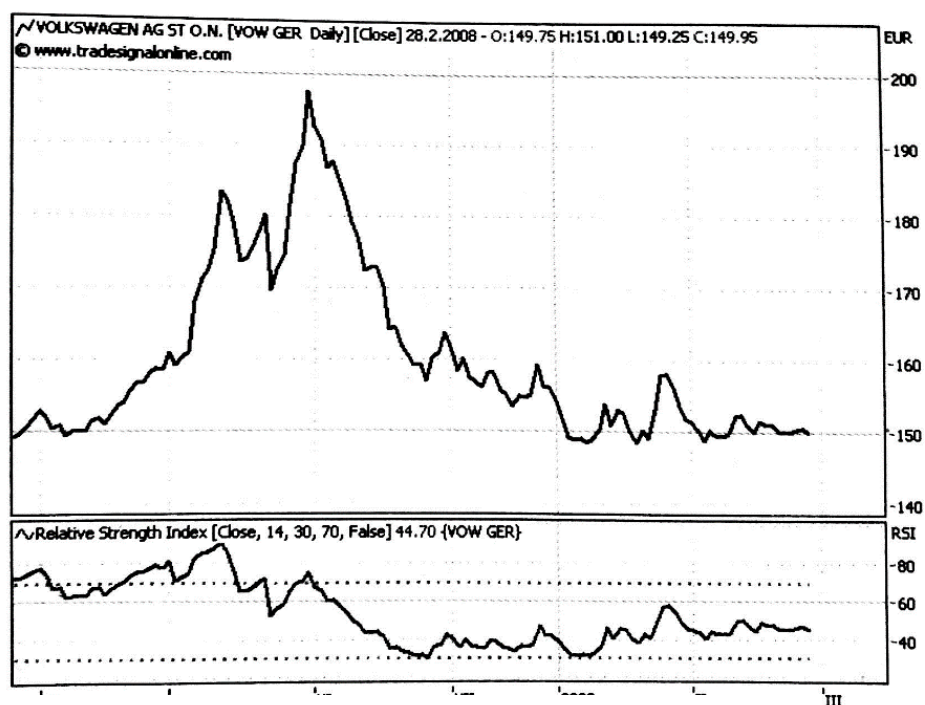
### **Rovnice 3: Relative strenght index**

$$RSI_t(n) = 100 - \left[ \frac{100}{1+RS(n)} \right] = 100 - \left[ \frac{100}{1+\frac{U(n)}{D(n)}} \right]$$

*Zdroj [16]*

Délka časového období  $n$  je opět na uvážení analytika, nicméně autorem J.W. Wilderem je doporučena délka 14 - denní. [16]

Zvláštní pozornost je třeba věnovat extrémním hodnotám, kdy se má za to, že hodnoty kolem 30 % značí předprodaný trh a kolem 70% trh překoupený. Dále je potřeba sledovat, zda se oscilátor vyvíjí stejným směrem, jako kurz akcie. Jestli-že oscilátor roste a kurz akcie klesá či stagnuje, lze očekávat růst a analogicky naopak. [25] [16] [3]

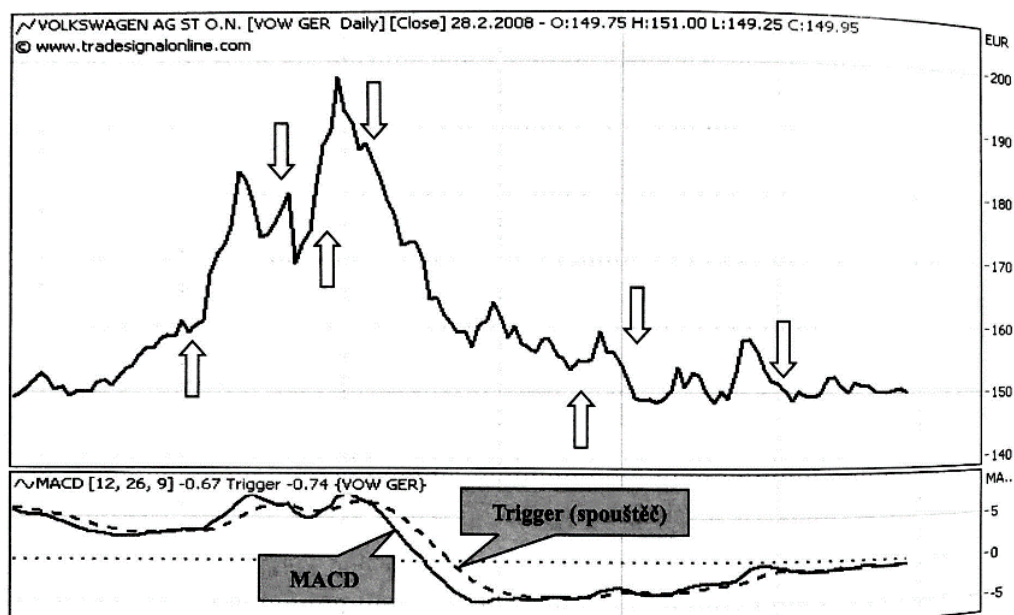


**Obrázek 18:** Vývoj Relative strenght index pro Volkswagen

*Zdroj [16]*

### *Moving Average Convergence Divergence (MACD)*

Oscilátor MACD je považován za jeden z nejpřesnějších technických indikátorů. Je definován jako rozdíl dlouhodobého exponenciálního klouzavého průměru (často se používá 25 - denní nebo 26 - denní) od krátkodobého indikátoru (obvykle 12 - denní). Tento oscilátor osciluje kolem středové linie procházející 0. Jako spouštěcí linie (trigger) k nákupní či prodejnímu signálům, zde slouží 9 - denní klouzavý průměr. Pokud MACD vzroste nad svoji spouštěcí linii, je generován nákupní signál, pokud poklesne pod ní, je naopak generován prodejní signál. Čím křivka MACD protne svoji spouštěcí linii dále od nulové přímky, tím je spouštěcí signál silnější. [25] [16] [3]



**Obrázek 19:** Vývoj MACD pro Volkswagen

*Zdroj: [16]*

#### Stochastik

Indikátor Stochastik vychází z předpokladu, že při rostoucím trendu je uzavírací kurz akcie v horní části denního rozpětí a při klesajícím trendu tomu je naopak. Vypočítat ho je možné podle následující formule:

#### Rovnice 4: Stochastik

$$\%K = \frac{\text{závěrečný kurz} - \text{minimální kurz}}{\text{maximální kurz} - \text{minimální kurz}} \cdot 100$$

*Zdroj [25]*

Znamená to tedy, jestliže se stochastik blíží k hodnotě 100, pak i jeho závěrečný kurz se blíží k maximálnímu kurzu v daném dni. Pokud se blíží směrem k 0, potom se i uzavírací kurz blíží směrem k minimálnímu kurzu za daný den. [25]

Dle výše uvedeného vzorce je konstruován takzvaný rychlý stochastik, který je velice volatilní a dává velké množství planých signálů. Proto se používá jeho klouzavý průměr (obvykle 3-denní), čímž se dostane takzvaný pomalý stochastik. Tento oscilátor již není tolik volatilní a negeneruje tak velké množství signálů. [25]

Pro identifikaci nákupních či prodejních signálů, je třeba sestrojít křivku %D, což není nic jiného, než klouzavý průměr stochastika (ať již rychlého nebo pomalého). Nákupní, respektive

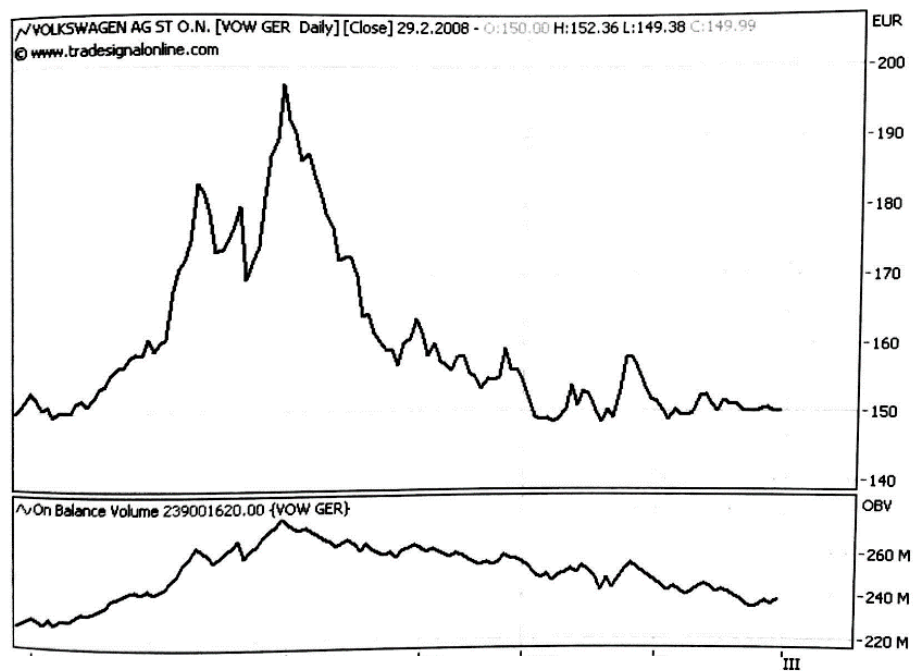
prodejní, signály poté jsou, pokud stochastik vzroste nad křivku %D, respektive klesne pod křivku %D. [25]

### **Objemové indikátory**

#### *On balance volume (OBV)*

Výpočet indikátoru bilance objemu je poměrně jednoduché. V první fázi se porovná uzavírací kurz dne s uzavíracím kurzem dne minulého. Pokud je vyšší, pak se k zobchodovanému objemu z aktuálního dne přiřadí kladné znaménko. Pokud je záporný pak se přiřadí znaménko záporné. Index OBV již pak je kumulativním součtem zobchodovaných objemů akcií zohledněným příslušným znaménkem. [25] [16]

Nákupním, respektive prodejním, signálem pak bude změna trendu indexu OBV. Pokud se mění z rostoucího na klesající, jedná se o prodejní signál, pokud z klesajícího na rostoucí, pak jde o signál nákupní. [25] [16]



**Obrázek 20:** Vývoj OBV pro Volkswagen

*Zdroj: [16]*



### *Money flow index (MFI)*

Smyslem tohoto indikátoru je porovnat mezi sebou kladné a záporné toky peněz mezi sebou. Index se získá dle následujícího vzorce:

#### **Rovnice 5: Money flow index**

$$MFI = 100 - \frac{100}{1 + MR}$$
$$MR = \frac{\textit{kladný tok peněz celkem}}{\textit{záporný tok peněz celkem}}$$

*Zdroj [25]*

Kde tok peněz získáme pomocí rozdílu mezi aktuálním průměrným denním kurzem a průměrný denním kurzem z minulého dne vynásobeným objemem aktuálního dne. Poté se za určité časové období sečtou kladné a záporné toky zvlášť a určí se poměr mezi nimi (MR). *MFI* se pak jen dopočte dle vzorce výše. [25] [16]

#### **1.2.3.2 ANALÝZA CELÉHO TRHU**

##### *Advance-decline line (A/DL)*

Indikátor *A/DL* je určený jako kumulativní počet rostoucích emisí, od kterých je kumulativně odečten počet klesajících emisí. Při tom je třeba sledovat ne tak hodnoty tohoto indexu, ale jeho trend. Dále všeobecně platí, že jestli indikátor *A/DL* roste a burzovní index klesá, lze očekávat, že nastoupí rostoucí trend. Analogicky naopak toto platí pro nástup klesajícího trendu. [25] [16]

Pokud bychom místo počtů rostoucích/klesajících emisí akcií sčítali/odčítali jejich objemy, dostali bychom indikátor *Upside-downside volume line*. [25] [16]

##### *Advance-decline ratio (A/DR)*

Indikátor *A/DR* získáme jako poměr počtu rostoucích emisí v daný den ku počtu klesajících emisí:

## Rovnice 6: Advance-decline ratio

$$ADR_t = \frac{\text{Počet rostoucích emisí}_t}{\text{Počet klesajících emisí}_t}$$

Zdroj [16]

Pro tento indikátor je opět klíčové správně nastavit horní a spodní hranici, kterou když indikátor překročí, bude značit signál k nákupu, respektive k prodeji. Tento indikátor bývá poměrně volatilní, proto se obvykle používá jeho klouzavý průměr (10-denní). [16]

Také v případě tohoto indikátoru, pokud počet rostoucích či klesajících emisí vyměníme za jejich zobchodovaný objem, dostaneme *Upside-downside volume ratio*. [16]

### *Indikátor New High-New Low Ratio*

Jak již název napovídá, pokud se vezme počet nově dosažených maxim v rámci všech emisí na trhu a vydělí se počtem všech dosažených minim za určité období (častokrát 52 týdnů), dostane se indikátor New High high-New low ratio. Tento indikátor se používá buď v této podobě, nebo v podobě jeho klouzavých průměrů. [16]

Důležité je tento indikátor interpretovat v souvislosti s burzovním indexem, kde divergence lokálních extrémů indikátoru vůči burzovnímu indexu znamená obrat trendu. Další možnou interpretací jeho hodnot jsou hranice 0,7, která signalizuje býčí trend. Při poklesu indikátoru pod 0,6 přichází varování před přechodem do medvědí fáze trhu, při které se indikátor pohybuje pod hodnotou 0,2. Při opětném vzestupu nad hodnotu 0,3 lze očekávat přechod do býčího trhu. [16]

## 1.3 Investiční omyly

V rámci kapitoly věnované investování na kapitálových trzích je vhodné čtenáře seznámit i s běžnými investičními omyly, které běžně drobní investoři dělají. Určitě každý již alespoň jedno v životě slyšel investiční přísloví: „nikdy nedávej všechna vejce do jednoho košíku“, které naráží na potřebnou diverzifikace investorova portfolia. Jmenovat a popsat všechny omyly by jistě vydalo na samostatnou diplomovou práci, proto je níže popsáno několik zajímavých z nich.

### **1.3.1 Máte nadměrné sebevědomí a přehnaně důvěřujete ve své investiční schopnosti**

Položme blíže neurčenému vzorku lidí otázku: „Máte s ostatními lidmi nadprůměrné dobré vztahy“? A poté otázku: „Jste nadprůměrný řidič“? Kolik procent lidí odpoví na tuto otázku kladně? 50 %? Méně? Nebo dokonce více?

Existuje mnoho studií, kdy celých 90 % respondentů odpovídá na tuto otázku kladně. Přitom z logiky věci, není možné, abychom měli 90 % nadprůměrných řidičů nebo 90 % lidí mělo nadprůměrné dobré vztahy s ostatními. Podobnou otázku, avšak z investičního světa položila společnost Montgomery Asset management investorům: „Jaké budou jejich výnosy?“ 74 % dotázaných odpovědělo, že podílové fondy, v nichž mají uložené prostředky, budou trvale porážet trh. Opět se jedná o nemožnou věc, aby 74 % investorů poráželo trh. Velké množství lidí, žene kupředu malá „jiskřička naděje“, že budou schopni porazit trh, a přitom jen velmi málo lidem se to opravdu podaří. K tomu je zapotřebí buď mít informace, které ostatní nemají (insider trading), což je protiprávní, nebo interpretovat informace lépe než ostatní.

Jaké z toho plyne poučení? Pokud si investor uvědomí svých velice omezených schopností předpovídat budoucnost a uvědomí si svou přirozenou lidskou tendenci k přehnanému sebevědomí, pak je na dobré cestě, stát se úspěšným investorem. [20]

### **1.3.2 Hledáte štěstí ve hvězdách**

Pokud se bavíme o hvězdách, nemyslíme v tomto kontextu obrovskou žhavou plasmovou kouli, která každou vteřinu spaluje nesčetně tun vodíku (či těžších prvků), nýbrž o ratingové agentuře Morningstar, která hvězdy přiděluje podílovým fondům. Rozsah, v kterém hvězdy přiděluje, je od jedné do pěti a měl by odrážet výkonost fondu po zohlednění rizika, v porovnání s porovnatelnými fondy či ETF. Smysl je takový, že nejlepších 10 % fondů dostane 5 hvězdiček a nejhorsích 10 % hvězdičku pouze jednu. Avšak může se investor na toto hodnocení spolehnout a založit na něm svoji investiční strategii?

A dle Morningstaru v roce 2004 pěťhvězdičkové fondy překonaly ty jednohvězdičkové o průměrně o 2,8 %. K této tezi je ovšem třeba dodat, že do pěti let tyto fondy spadly na průměrný rating 3,1 hvězdičky. Této tezi ani moc nepomůže studie, kdy pěťhvězdičkové fondy překonají pouze 59 % aktivně řízených fondů. Tyto aktivně řízené fondy, avšak v průměru zaostávají o více, než 1,5 % p.a. za svým benchmarkem. Roku 2007 se pěťhvězdičkovým fondům dokonce tak nedařilo, že zaostaly za fondy jednohvězdičkovými. [20]

### **1.3.3 Myslíte si, že kvalita diverzifikace je určena počtem cenných papírů**

Někteří investoři se drží rady „nedávej všechna vejce do jednoho koše“ a tedy drží v portfoliu velké množství titulů. A čím více tím lépe, tím je jejich portfolio více diverzifikované, méně rizikové a nic jim nehrozí. Bohužel diverzifikace se neodvívá až tolik od počtu aktiv držených v portfoliu, nýbrž od korelace mezi sebou. Tedy ač bude portfolio sestavené z velkého počtu různých akcií, avšak pokud budou spolu korelované, tak o diverzifikovaném portfoliu se moc hovořit nedá. [20]

Krátká poznámka nakonec: Pokud jste právě zjistili, že nemáte zrovna diverzifikované portfolio, tak nemusíte být smutní, v dobách krize obvykle padá cena všech aktiv, i sebelépe diverzifikovaných portfolií.

### **1.3.4 Věříte v „horkou ruku“ manažerů fondu**

Představme si 2 srovnatelné fondy. Jeden již třetím rokem se zařadil mezi fondy s nadprůměrným výnosem, druhý zase naopak třetím rokem mezi ty podprůměrné. U jakého fondu bude větší pravděpodobnost, že se čtvrtý rok zařadí mezi ty nadprůměrné? U žádného. Oba dva fondy se tam zařadí s pravděpodobností 0,5. Viz třeba takový Wall Street Fund, jehož již dávno zapomenutý portfolio manažer David Baker přinesl svým investorům nevídané zisky. V letech 1970-1980 byl nejvýkonnějším fondem investující do amerických akcií. Celých 10 let, kdy překonal i takové hráče, jako třeba Peter Lynch. Bohužel v následující dekádě, ztratil celých 73 % a skončil jako úplně poslední (index S&P rostl průměrně 17,5 % ročně). [20]

Jaké z toho plyne ponaučení? Fakt, že některý fond má nadprůměrné zisky, ještě neznamená, že ho řídí génius či věštec. Dost možná měl portfolio manažer prostě jen štěstí. Přece jenom existují tisíce fondů a pravděpodobnost, že alespoň jeden z nich bude delší sérii let nadprůměrný, není vůbec malá.

## 2 DATA ANALYTICS & DATA SCIENCE

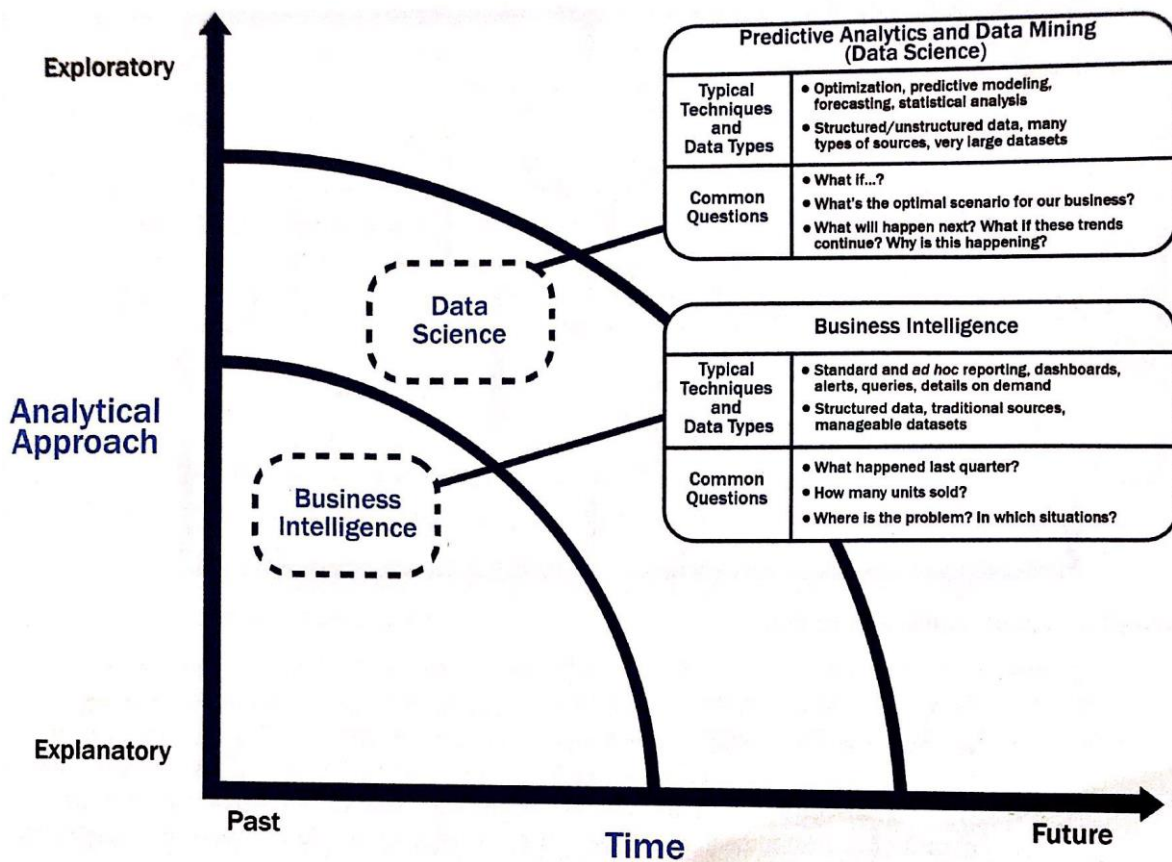
Tato kapitola je věnována tématu, které je v poslední době velmi často skloňované, data analytics a data science. Důvodem je, že se ukládá a uchovává čím dál více dat, které je možné analyzovat, a zároveň roste poptávka ze strany společností, které je chtějí využít jinak, než aby jim jen ležela v „data warehouse“ (datovém skladišti).

Každý, kdo má účet na Facebooku, či používá Google se s výsledky data analytics již setkal. Reklamy na postranních lištách se nezobrazují náhodně nýbrž na základě algoritmu (ať již Facebooku nebo Googlu), do něhož vstupuje řada proměnných (ať již osobní údaje nebo historie stránek) a který má za úkol nabídnout výrobek, který má největší pravděpodobnost zaujmout uživatele. Proto kupříkladu se z pravidla mužům kolem 25 moc nebudou zobrazovat rtěnky či kyčelní protězy a naopak ženám kolem 60 doplňky do auta či skoky s padákem.

### 2.1 Co to je?

Když se řekne Data science, co si teda vlastně pod tím lze představit? Jak již bylo naznačeno v úvodu, hlavním smyslem data science je pomocí dat, obvykle se jedná o velké množství dat („big data“), a analytických nástrojů získat doporučení či postřehy („insights“). Tyto postřehy či doporučení pak v ideálním případě ústí v rozhodnutí, zde se používá termín „data driven decisions“ – rozhodnutí učiněné na základě analyzování dostupných dat.

Jak někteří možná ví, existuje disciplína s podobným zaměřením: Business intelligence. Cílem je též zásobit lidi, kteří rozhodují (rozhodovatele), různými postřehy či doporučením. Avšak jak je zřejmé z obrázku 21, je zde jeden zásadní rozdíl. [2]



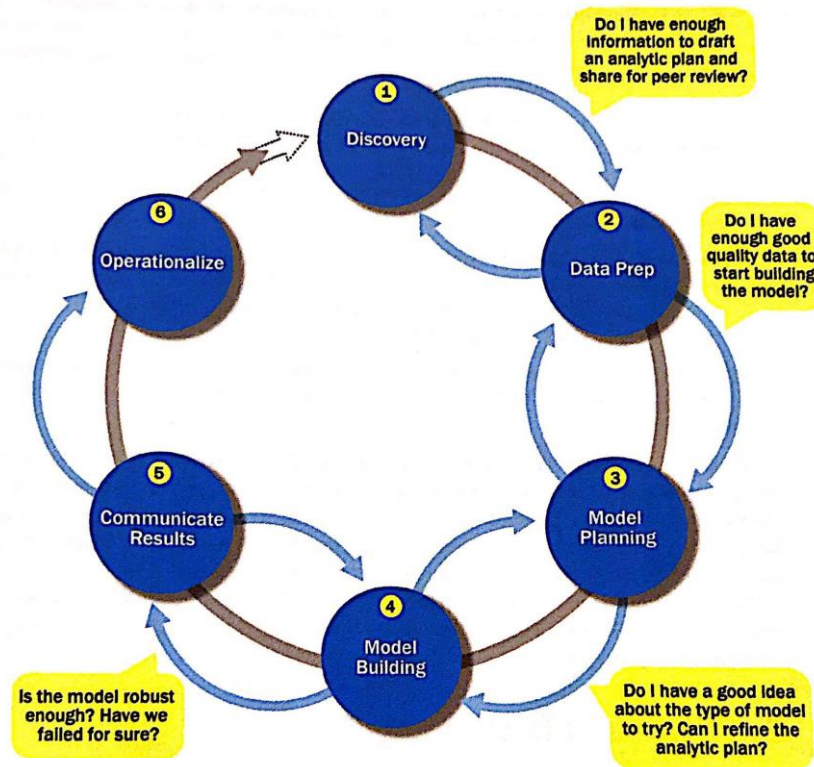
Obrázek 21: BI vs Data Science

Zdroj: [2]

Hlavním rozdílem je, kam tyto disciplíny hledí. Zatímco Business Intelligence hledí na přítomnost či minulost, Data science hledí spíše dopředu, na budoucnost. Z toho vyplývá i forma výstupů, kde v případě Business intelligence jde z pravidla o statické reporty či dashboardy. Odpovídá převážně na otázku: kdy a kde, se něco stalo? Data science již jde o krok dále a výstupy jsou často prediktivní modely, předpovědi či statistická analýza. Otázky, které si pokládá, bývají: Jak? Proč? Co když...? Za zmínku určitě stojí také data, z kterých tyto disciplíny čerpají. V případě Business intelligence jde obvykle o strukturovaná data pocházející z jednoho zdroje, a tyto data dále agreguje či seskupuje („grupuje“). Data science jde opět o krok dále a využívá vedle strukturovaných dat i data nestrukturovaná a kombinuje několik zdrojů dohromady, aby výsledek byl komplexní. [2]

## 2.2 Životní cyklus

Životní cyklus projektu pro Data analytics je možné rozdělit do šesti fází, které je možné vidět na obrázku 22, a budou blíže popsány v následujících podkapitolách. Projekt se ale může v jeden časový okamžik, nacházet ve více fázích najednou. Dále je možné, že se projekt vrátí do předchozí fáze (předchozích fází), pokud dojde k objevení nových informací či poznatků. Před popisem jednotlivých fází životního cyklu projektu, je třeba definovat klíčové role pro úspěšný projekt. [2]



Obrázek 22: Životní cyklus projektu

Zdroj [2]

### 2.2.1 Klíčové role

**Business user = Uživatel**

Člověk, který je expertem v dané oblasti a také nejčastěji benefituje z výstupu daného projektu, má zde roli poradce projektového týmu, aby jejich výsledky byly opravdovým přínosem v dané problematice. [2]

**Project sponsor = Sponzor projektu**

Tento člověk je zodpovědný za vznik projektu. Definiuje klíčový problém, priority projektu a očekávaný výsledek. Obstarává budget a pak měří dopad výsledků projektu. [2]

### **Project manager = Projektový manažer**

Projektový manažer zabezpečuje, aby byly klíčové body v projektu naplněny včas a v požadované kvalitě. [2]

### **Business Intelligence analyst = BI analytik**

Analytik, který je zodpovědný za přípravu dashboardů a reportů. Jeho expertíza je v oblasti interpretace dat a základních metrik v kontextu reportingu. [2]

### **Database administrator = Databázový administrátor**

Spravuje a konfiguruje prostředí databáze, aby vyhovovalo požadavkům analytiků. Vedle toho jim zajišťuje přístupy do databáze. [2]

### **Data engineer = Datový inženýr**

Expert, který pomocí SQL dotazů pomáhá dostat data z databází do požadovaného datového modelu. Úzce spolupracuje s datovými vědci, kterým pomáhá ztvárnit data do požadované formy pro analýzy. [2]

### **Data scientist = datový vědec**

Expert v oblasti analýzy dat, modelování a aplikování správných analytických postupů na problémy businessu. Zodpovídá za naplnění všech analytických cílů. [2]

## **2.2.2 Discovery**

V první fázi je klíčové, aby se tým datových vědců seznámil s problémem, který budou řešit. Důležité je, aby ohraničili problém a definovali, jaké jsou hlavní výzvy a jaké dílčí aktivity spadají do tohoto problému a které již ne („out of scope“). Dalším důležitým krokem je identifikace a následná komunikace s klíčovými lidmi, kterých se problém týká ohledně jejich očekávání (včetně sponzora projektu). Vedle komunikace je také důležitá průběžná spolupráce s těmito lidmi. V závěru této fáze, po interview se sponzorem projektu a klíčovými lidmi, je již možné formulovat prvotní hypotézy a určit potenciální datový zdroj. [2]

## **2.2.3 Data preparation**

V rámci této fáze analytici nejdříve objevují, jaká data mají k dispozici, v jaké jsou kvalitě a následně je upravují do potřebné podoby. Příprava dat je paradoxně časově nejnáročnější fáze, běžně týmy zde stráví i 50 % z celkového času. Tato fáze je velice důležitá, jelikož platí zlaté



pravidlo „garbage in, garbage out“, neboli pokud vstupem do našeho modelu, budou špatná, nevyčištěná data, pak nelze očekávat relevantní výsledky. [2]

Aby nedocházelo k zásahům ze strany analytického týmu do databází živého provozu, vytváří se takzvaný „analytic sandbox“. Tedy jakýsi pracovní prostor, kde mohou analytici zkoumat a upravovat data, ale nijak nezasahují do živého provozu, jelikož se jedná o obraz reálných databází. [2]

#### **2.2.4 Model planning**

V průběhu této fáze analytici identifikují vhodné kandidáty na model, který použijí v analýze a zároveň proměnné, které budou do něj vstupovat. Zde se odkazují na počáteční hypotézy, aby vybraný model byl optimální na jejich testování a naplnil rámec analytického problému. [2]

#### **2.2.5 Model building**

Je potřeba rozdělit množinu dat na trénovací množinu dat, na které se model „učí“, a na testovací množinu, která slouží pro validaci modelu. Zároveň je klíčové zajistit, aby obě tyto množiny byly dostatečně robustní, jinak výsledky mohou být značně zkreslené. [2]

Je celkem běžné, že v praxi dochází k iteraci mezi fází plánování modelu a stavění modelu, než dojde k sestavění finálního modelu. Ač jsou modelovací techniky poměrně složité a komplexní, fáze plánování a stavění modelu jsou ve výsledku výrazně kratší v porovnání s fázemi přípravy dat či prezentace a interpretace výsledků. [2]

#### **2.2.6 Communicate results**

Po přípravě modelu je třeba porovnat výsledky modelování s kritérii, pro úspěch či neúspěch, které byly definovány na začátku. Tým se rozhodne, jak nejlépe interpretovat výsledky analýzy ostatním týmům a ostatním zainteresovaným stranám. Musí vzít v úvahu všechna upozornění, předpoklady a limitace modelu. Vzhledem k tomu, že obvykle audience je napříč celou firmou, musí zvolit vhodnou formu, aby to bylo srozumitelné pro všechny. Hlavním cílem je demonstrovat klíčová zjištění a předání hlavní zprávy. [2]

#### **2.2.7 Operationalize**

Ve finální fázi projektu tým na základě výstupů z modulu, připraví pilotní projekt. Než implementuje projekt do celého ekosystému organizace, projekt nasadí nejdříve pouze v části organizace, která bude pod průběžnou kontrolou. [2]

## 2.3 Metody

Pro účel analyzování a modelování dat existuje velká řada metod a každou je vhodné použít v jiném případě. Níže budou proto popsány ty nejběžněji používané.

### 2.3.1 Lineární regrese

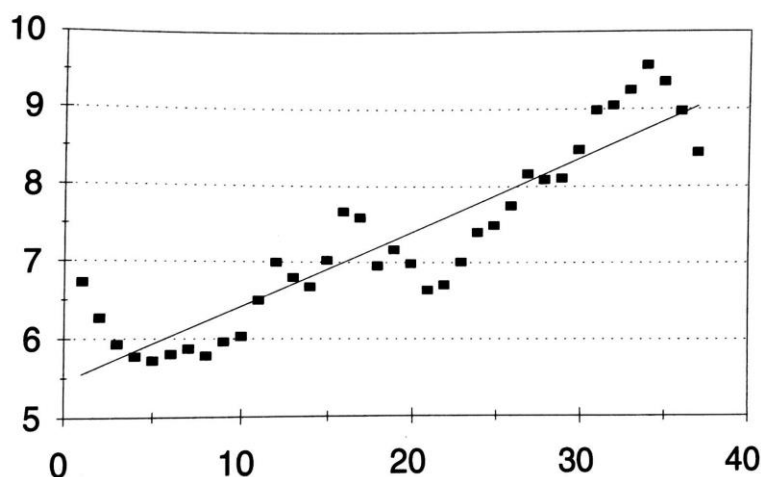
Lineární regresi je vhodné použít, pokud chceme vysvětlit číselnou proměnnou pomocí jedné nebo více z pravidla také číselných proměnných, nicméně je možné použít i kategorické proměnné či kombinaci s číselnými proměnnými. Pokud se budeme pohybovat v  $n$  rozměrném prostoru a budeme brát v úvahu  $m$  proměnných, pak by matematický zápis mohl vypadat třeba jako vzorec 2.1. [9] [10]

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1m} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2m} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ x_{n1} & x_{n2} & x_{n3} & \cdots & x_{nm} \end{bmatrix} \times \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \beta_m \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_n \end{bmatrix} \quad [2.1]$$

*Zdroj:[9]*

Zde je  $(y_1 \dots y_n)$  právě vektor vysvětlované proměnné,  $(x_{1,1} \dots x_{n,m})$  je matice  $m$  proměnných v rámci  $n$ -rozměrného prostoru,  $(\beta_1 \dots \beta_m)$  je vektor koeficientů pro jednotlivé proměnné a  $(\varepsilon_1 \dots \varepsilon_n)$  je vektor náhodných odchylek lineárního modelu od reálných dat, které chceme pomocí něho vysvětlit. Tento model by bylo vhodné ještě obohatit o nenulový vektor  $(\alpha_1 \dots \alpha_n)$ , který by se k celému modelu přičetl. [9]

Na obrázku 23, si můžeme graficky představit lineární regresi. Zde se snažíme proložit naměřená data přímkou (či jinou křivkou) aby data co nejlépe vystihovala a s co nejmenší možnou chybou. [10]



**Obrázek 23:** Grafická ilustrace lineární regrese

*Zdroj: [10]*

V praxi se nejčastěji používá pro lineární regresi metoda nejmenších čtverců. Nutná, nikoliv postačující podmínkou této metody je, že suma všech odchylek musí být rovna nule. Postačující podmínkou je, že suma čtverců všech odchylek je minimální. Další podmínkou je, že odchylky mají konstantní a konečný rozptyl. Podmíněný rozptyl  $D(y/x) = \sigma^2$  je také konstantní a konečný. Data jsou tedy homoskedastická. Odchylky jsou nekorelované a nezávislé a mají normální rozdělení  $N(0, \sigma^2)$ . [9]

Po nalezení koeficientů  $\beta$  je třeba otestovat, jestli jsou tyto koeficienty statisticky významné. To znamená, že pokud bychom je nahradili nulou, model by výrazným způsobem ztratil svoji vypovídající schopnost. Toto provedeme pomocí testovacího kritéria:

$$F = \frac{(\mathbf{b} - \beta_0)^T \mathbf{X}^T \mathbf{X} (\mathbf{b} - \beta_0)}{m \hat{\sigma}^2} \quad [2.2]$$

Kde  $\beta_0 = 0$ ,  $\mathbf{X}$  je matice proměnných a  $\mathbf{b}$  je vektor koeficientů. Toto testovací kritérium má Fisherovo rozdělení s  $m$  a  $(n-m)$  stupni volnosti. [9]

Poslední otázkou je posouzení kvality modelu. Tedy jak hodně náš model vystihuje vysvětlovanou proměnnou pomocí vysvětlujících proměnných. Pro tento účel se používá index determinace, který je definován jako podíl variability vysvětlené modelem a celkové variability, pohybující se na intervalu  $<0;1>$ . Nicméně neexistuje žádná hranice, od které by byl model "dostatečně vypovídající". [9] [10]

Index determinace můžeme vypočít pomocí vzorce [2.3]

### Index determinace

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad [2.3]$$

Zdroj: [9]

Kde  $\hat{y}$  je predikovaná hodnota dle modelu,  $\bar{y}$  je průměrná hodnota vysvětlované proměnné a  $y$  je hodnota vysvětlované proměnné. [9]

### 2.3.2 Logistická regrese

Pokud chceme modelovat spojitou proměnnou, tak jak bylo řečeno v předchozí kapitole, je vhodné použít například lineární regresi. Co když ale chceme modelovat kategorickou proměnnou? Pro tento případ tu máme logistickou regresi. Konkrétně se nejčastěji používá pro modelování alternativní proměnné. Vysvětlující proměnné pak mohou opět být jak spojité (či diskrétní), tak kategoriální. [9]

Vezměme v úvahu případ, že chceme modelovat proměnnou, která nabývá pouze hodnot 1 a 0. Příkladem může být například bankovní scoringový model, který modeluje, jestli dlužník spadne do defaultu nebo ne. Logistickou regresí pak modelujeme pravděpodobnost  $p$ , že modelovaná hodnota nabyde 1 a opačnou pravděpodobnost  $(1-p)$ , že nabyde 0. Tato pravděpodobnost je obsažena v logitové transformaci: [9]

$$L_{(1)} = \text{logit}(p) = \ln \frac{p}{1-p} \quad [2.4]$$

Po úpravě pak dostaneme:

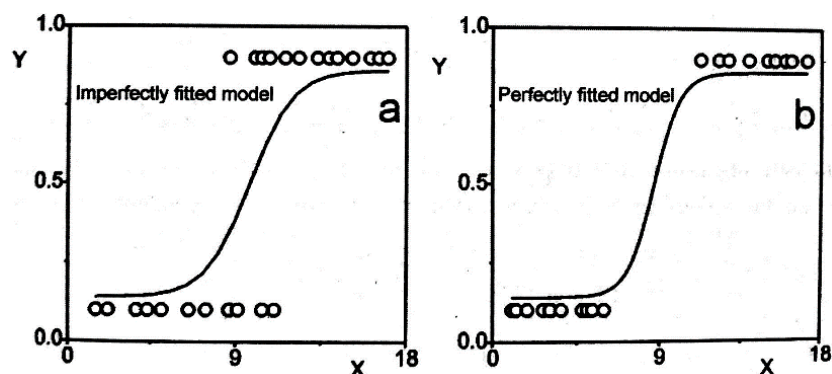
$$p = \frac{e^{L_{(1)}}}{1+e^{L_{(1)}}} \quad [2.5]$$

$L_{(1)}$  pak dostaneme jako model lineární regrese:

$$L_{(1)} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m \quad [2.6]$$

Kde  $\beta_0$  je konstanta a  $\beta_1$  až  $\beta_m$  jsou koeficienty pro  $m$  proměnných  $x$ .

Na obrázku 24 pak je možné vidět logitovou funkci nejdříve pro špatně nastavený model, který má slabé predikční schopnosti, a poté pro dobře nastavený model. [9]



Obrázek 24: Logitová funkce

Zdroj: [9]

### 2.3.3 Rozhodovací stromy

Rozhodovací stromy jsou jedním z nejoblíbenějších algoritmů strojového učení. Důvod, proč jsou tak oblíbené je, že pomocí vizualizace jsou snadno a rychle pochopitelné. Rozhodovací stromy analyzují set dat tak, že definují set otázek/pravidel, které determinují jejich třídu. Rozhodovací stromy je tedy vhodné použít, pokud chceme predikovat závislou spojitou či kategoriální proměnnou. V případě spojitě proměnné se pak mluví o regresních stromech, pro kategoriální jsou to stromy klasifikační. [18]

Například představme si, že máme množinu zvířat, ke které máme základní charakteristiky jako váhu, počet nohou, velikost a třídu. A nyní chceme určit, podle jakých pravidel by šly jednotlivá zvířata na základě informací, které o nich víme, zařadit do určité třídy. Jako příklad třídícího pravidla můžeme uvést hmotnost zvířete pod jeden gram. Zvířata splňující toto kritérium bychom například třídili jako hmyz. Nicméně na další pravidla bychom již potřebovali potřebná data a pomoc algoritmu.

V praxi se nejčastěji používá **C&RT** (classification and regression trees), **CHAID** (Chi squared Automatic Interaction Detector), **QUEST** (Quick, Unbiased, Efficient Statistical Trees) a **C5.0**. Hlavní rozdíl mezi těmito metodami je v kritériu, podle kterého se dále větví atributy, v tom jestli zvládají spojitý vstup či spojitý výstup a ve způsobu větvení. Rozdíly jsou blíže zobrazeny v tabulce 3. [18] [13]

**Tabulka 3:** Porovnání rozhodovacích stromů

<b>Vlastnosti</b>	<b>C&amp;RT</b>	<b>C 5.0</b>	<b>CHAID</b>	<b>QUEST</b>
<b>Dělení</b>	binární	vícenásobné	vícenásobné	binární
<b>Zvládá spojitý výstup?</b>	ano	ne	ano	ne, náhrady
<b>Zvládá spojitě vstupy?</b>	ano	ano	ne	ano
<b>Kritérium výběru atributu</b>	Giniho index	poměrný informační zisk	Chi $X^2$ test	statistické
<b>Rychlost</b>	nižší	vyšší	střední	střední/nižší
<b>Chybějící údaje</b>	ne, náhrady	ne, náhrady	ano, jako kategorie	ne, náhrady

*Zdroj: upraveno dle [13]*

### 2.3.4 Simulace Monte Carlo

Simulace Monte Carlo je nejčastěji využívána v případech, že je velmi obtížné či nemožné použít klasické analytické výpočetní metody. Principem této metody je, že se nasimuluje chování matematických nebo fyzikálních systémů. Metoda umožňuje modelovat každý proces, na který mají vliv náhodné faktory a její základní myšlenka je založena na spojitosti a vztahu mezi pravděpodobnostními charakteristikami různých náhodných procesů. Příprava takovéto simulace bývá poměrně časově náročná a je potřeba velký počet průběhů pro analýzu přesnosti a citlivosti. Na druhou stranu je poměrně oblíbená, jelikož záběr problémů, které pomocí ní lze vyřešit, je velice široký, a je jednoduchá na vysvětlení a konceptualizaci. Také je to jediná metoda, která je schopna modelovat velkou řadu komplexních stochastických systémů. [10] [6] [21] [8]

Metodu můžeme demonstrovat různými způsoby. Například při odhadu velikosti čtverce pomocí umístění náhodných bodů. Avšak oblíbeným příkladem na demonstraci simulace Monte Carlo je výpočet Ludolfova čísla. Představme si, že vůbec netušíme, kolik Ludolfovo číslo je, a počítač, ani jiná výpočetní technika, toto číslo nezná. Sestrojíme tedy čtverec o délce strany  $2r$  a do něj vepíšme kružnici s poloměrem  $r$ . Nyní budeme zcela náhodně umísťovat body v rámci celého čtverce a určíme poměr, kolik těchto bodů leží uvnitř vepsané kružnice vůči všem bodům. Tento poměr je pro nás klíčový, jelikož v sobě skrývá poměr obsahu čtverce a obsahu kružnice, a z toho poměru pak můžeme dle úpravy rovnice 8 získat Ludolfovo číslo, respektive jeho jednu čtvrtinu. Čím více simulací provedeme, tím toto číslo bude blíže reálnému číslu  $\pi$ . [6] [8]

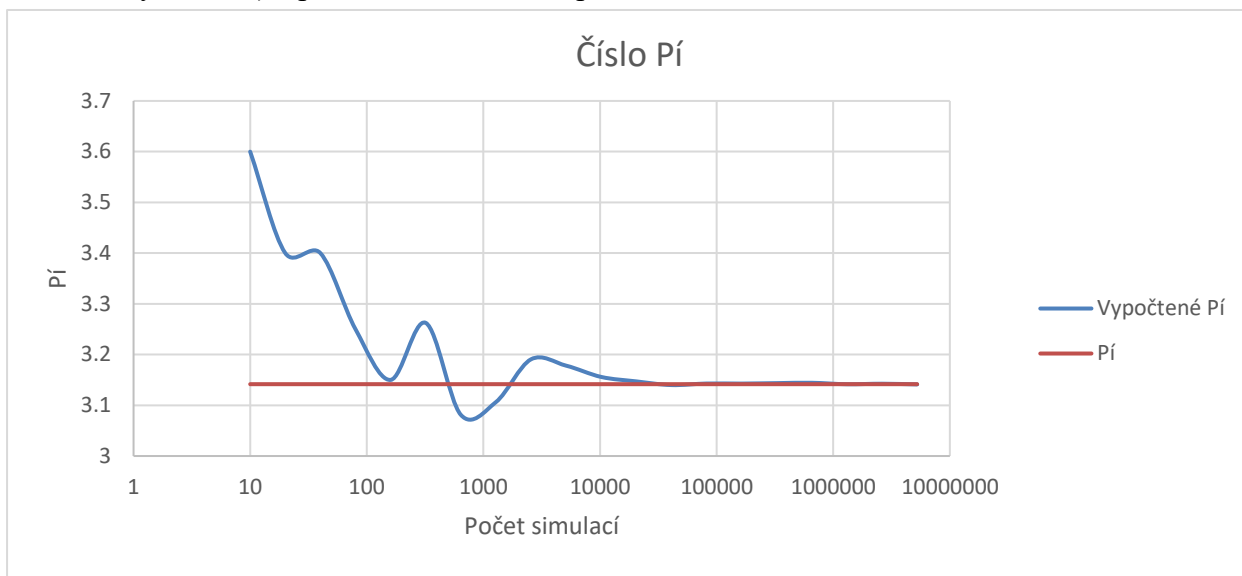
## Poměr obsahu kružnice a čtverce

$$\frac{\pi r^2}{2r \cdot 2r} = \frac{\pi}{4} \quad [2.7]$$

Jelikož dělat velký počet náhodných bodů na papíře, a pak počítat jejich počet, by bylo velmi náročné, zkusíme to udělat pomocí výpočetní techniky. Jako nástroj je zvolen programovací jazyk Python. Poloměr kružnice  $r$  zvolme 0,5. Nyní je potřeba sestavit algoritmus, který náhodně vygeneruje souřadnice bodu v dvojrozměrném prostoru z intervalu  $\langle 0;1 \rangle$  a zkontroluje, jestli tento bod leží uvnitř vepsané kružnice do čtverce o poloměru 0,5. Pro tento účel změříme vzdálenost mezi bodem a středem kružnice  $[0,5;0,5]$ . Jestli bude tato vzdálenost menší nebo rovna, než je poloměr kružnice (0,5), pak tento bod leží uvnitř kružnice. Pro měření vzdálenosti si pomůžeme Pythagorovou větou a následovně sledujeme poměr mezi všemi body a body uvnitř kružnice, který by měl být přibližně jedna čtvrtina  $\pi$ . Kód tohoto algoritmu může vypadat například takto: [6]

```
import numpy as np
from random import *
import math
#Variables declared
scenarios = 10
r = 0.5
mypi = np.zeros((25,2))
cs = 0
#Cycle to try more and more scenarios with stopping condition 10 mil. scenarios
while scenarios < 10000000:
    points = np.zeros((scenarios,2))
    inside = 0
    #Cycle for generating random points and deciding if they are inside the circle
    for i in range (0, scenarios):
        for j in range (0, 2):
            points[i,j]=random()
            if math.sqrt(math.pow(points[i,0] - r,2) + math.pow(points[i,1] - r,2)) <= r:
                inside += 1
    #Pi calculation and inserting into the array
    pi = inside/scenarios*4
    mypi [cs,0] = scenarios
    mypi [cs,1] = pi
    cs += 1
    scenarios = scenarios*2
print(mypi)
```

Dále můžeme pozorovat vztah mezi zvětšujícím se počtem scénářů (vygenerování náhodných bodů) a přesností určení čísla pí. Tento vztah můžeme vidět na obrázku 25.



**Obrázek 25:** Výpočet Pí v závislosti na počtu simulací

*Zdroj: vlastní zpracování*

## 2.4 Python

Python je nejpoužívanějším a nejoblíbenějším programovacím jazykem v oblasti data analytics & data science. Pokud se podíváme na oficiální stránky Pythonu, najdeme zde výčet výhod, které Python přináší. Developéři zde uvádí, že je Python rychlý, silný, že ho lze spustit všude, komunikuje v rámci ostatních vývojových prostředí, je uživatelsky přívětivý a v neposlední řadě jednoduchý na naučení. Python obsahuje nástroje jak pro datovou analýzu, tak pro strojové učení, umělou inteligenci, webové i softwarové vývojáře. Díky své univerzálnosti je také velmi obratný, jelikož obsahuje velkou řadu funkcí, a zároveň ho lze integrovat s velkou řadou prostředí. [14]

V Python brožuře se můžeme dočíst, že Python využívají společnosti, jako Google, Youtube, Mozilla, Quantax a další. Tyto firmy zde také uvádějí, proč si vybraly právě Python. Téměř všechny tyto články spojují dvě slova: speed a agility. [15]

Python je opravdu rychlý, dokáže miliony operací za vteřinu a ostatní jazyky, jako Visual Basic, Java či C++, mu nestačí. Díky své univerzálnosti je také velmi obratný, jelikož obsahuje velkou řadu funkcí a zároveň ho lze integrovat s velkou řadou prostředí. [14]

### 2.4.1 Krátký tutoriál

Tato kapitola je věnována krátkému přiblížení programovacího jazyka Python. Obsahuje základní prvky syntaxe, objektů, funkcí a metod. [19]



#### 2.4.1.1 SYNTAXE

Kód Pythonu je velmi jednoduchý na čtení a na pochopení. Stejně jako každý programovací jazyk, kód je formátován do bloků. Na rozdíl od ostatních používá méně jednoduchých a složených závorek. Blok lze identifikovat pomocí dvojtečky na konci řádku kódu. Jedinou nevýhodou je, že každý řádek kódu v bloku musí mít stejné odsazení, jinak se objeví syntaxová chybová hláška. [19]

Komentování lze v prostředí Pythonu učinit dvěma způsoby. Buď pomocí „hashtagu“ (#), který značí, že vše za ním, až do konce řádku, je komentář. Dalším způsobem jsou trojitě uvozovky, které signalizují, že vše až do ukončení (opět trojitými uvozovkami) je komentář. Tímto způsobem je možné napsat komentář přes více řádků. [19]

Pokud bychom chtěli napsat více deklarování na jeden řádek, tak je to možné udělat pomocí oddělení středníkem. Naopak pokud bychom chtěli jeden řádek rozdělit na více (kvůli lepší čtivosti kódu), pak je toto možné udělat pomocí zpětného lomítka (\) na konci zalamovaného řádku. [19]

#### 2.4.1.2 ČÍSLA

##### *Integer*

Python se odkazuje na integer jako int. Jedná se o celá čísla v rozsahu  $[-2^{31}, 2^{31} - 1]$ . Nejčastější operace jsou sčítání, odčítání, násobení. Problém může nastat u dělení, jelikož dělením dvou celých čísel nemusí vzniknout celé číslo. Výhodou integerů je menší nárok na paměť, než třeba u reálných čísel. [19]

##### *Reálná čísla*

V Pythonu je na ně odkazováno jako na float. Reálná čísla už jsou desetinná čísla s přibližnou přesností na 16 desetinných míst a jsou v rozsahu  $[-10^{308}, 10^{308}]$ . S reálnými čísly je možné provádět velkou řadu i pokročilých matematických operací. Pro tyto operace je vhodné naimportovat modul Math pomocí příkazu `import Math`. [19]

## *Boolean*

Pro úplnost je vhodné zmínit Boolevu algebru. Pokud je číslo nadefinováno jako boolean, pak může nabývat dvou hodnot: True nebo False (ekvivalent 1 a 0). S těmito čísly je pak možné provádět operace jako konjunkce (AND), disjunkce (OR), negace (NOT) a další. [19]

### **2.4.1.3 KONTEJNERY OBJEKTŮ**

Programovací jazyky mají jednu velmi užitečnou schopnost, a to dělat opakované úkony za velmi krátký čas. Pro tento účel jsou vybaveny kontejnery objektů, často nazývané jako pole, které sdružují velký počet objektů stejného typu, a na které se lze odkazovat pomocí indexace. Matematické vektory pak mohou být zkonstruovány jako jednorozměrná pole, matice pak jako dvojrozměrná pole. Poměrně překvapující je, že v základní jazykové sadě Pythonu pole nenalezneme. Python využívá koncept více obecných kontejnerů, než jsou pole a setkáme se zde s prvky jako: list, tuple, string a dictionary. Pokud bychom ale trvali na využití konceptu polí, pak je možné je využít po nainstalování knihovny Numpy. [19]

## *List*

List je podobný poli, avšak není zde nutná podmínka, aby objekty byly stejného typu. Jeden list tedy může naráz obsahovat celá čísla, reálná čísla, slova a další typy. Je definován pomocí objektů oddělených čárkami a uzavřených do hranatých závorek. S listy může být dále zacházeno jako se samostatnými objekty a lze s nimi vykonávat matematické operace. [19]

## *Tuple*

Tuple je velmi podobný listu. V zápisu se liší pouze užitím kulatých závorek místo hranatých. Ve významu je zde však jeden klíčový rozdíl. Po deklarování již není možné změnit Tuple, je tedy neměnný v prostoru a čase. [19]

## *String*

String je vlastně textové pole, kdy jednotlivé znaky jsou zde objekty, které se neoddělují čárkami, respektive ničím. String je ohraničený uvozovkami, buď jednoduchými, nebo dvojíty, ale nikdy ne kombinací obou. Indexování a procházení stringu funguje na stejném principu, jako tomu je u listů či polí. [19]

## Dictionary

Dictionary, neboli slovník, je podobný Tuple, jelikož je také neměnný. Rozdílem je, že na jednotlivé objekty se zde neodkazuje pomocí indexace, ale nýbrž pomocí takzvaných klíčových slov, které uvedeme při deklarování. Slovníky jsou ohraničené složenými závorkami, jednotlivé objekty jsou opět odděleny čárkami a klíčová slova od objektů jsou odděleny dvojtečkou. [19]

### 2.4.1.4 IF-ELSE PODMÍNKY

Pomocí podmínky je nejjednodušší způsob, jak změnit chování algoritmu. Jsou obsaženy v každém programovacím jazyku. V Pythonu je syntaxe opět jednoduchá: if, poté podmínka, která generuje odpověď True nebo False, a nakonec řádku dvojtečka. Následující blok kódu, který definuje, co se má vykonat, je-li splněna podmínka, musí být odsazen. Za blokem může (ale nemusí) následovat doplňující else:, které definuje, co se má vykonat, pokud podmínka splněna není, nebo případně elif, kterým se definuje další podmínka (podmínky). [19]

### 2.4.1.5 CYKLY

Cykly je vhodné použít, pokud potřebujeme napsat nějaký iterační proces. Tento proces pak může být ukončen buď podmínkou, určitým počtem opakování, nebo projitím všech prvků určitého objektu. [19]

Příkladem může být postupné procházení všech znaků určitého slova a poté další analyzování těchto znaků. Nebo projití všech prvků listu a poté jejich analýza. Syntaxe je pro změnu opět jednoduchá: for iterator in iterable:. Za slovo iterable si pak dosadíme nějakou množinu objektů (list, pole, množina čísel...) a iterator nějaký kvantifikátor, například proměnnou. [19]

Vedle for cyklů existují while cykly, které se opakují pořád dokola, dokud je splněna určitá podmínka. Jejich syntaxe je: While condition:. Za slovo condition dosadíme podmínku. [19]

## 2.4.2 Numpy

Numpy je knihovna vystavěná na Pythonu a je vhodná pro výpočty v oblasti data science. Její velkou výhodou je, že umožňuje práci s  $n$ -rozměrnými poli. Tyto pole představují rychlou a flexibilní práci s velkým množstvím dat. Vedle práce s poli nabízí také řadu matematických funkcí, jako metody lineární algebry či Fourierovu transformaci. [19]

### 2.4.3 Pandas

Co to Pandas je? Jak by název možná mohl napovídat, tak se nejedná o žádný nástroj pro pandy, nýbrž opět o analytickou knihovnu vystavěnou na programovacím jazyku Python. Je to robustní sada nástrojů umožňující analýzu, filtrování, manipulaci, agregování, spojování, transformaci a čištění dat. Obsahuje také řadu funkcí sloužících k modelování časových řad. Tuto sadu lze přirovnat k Excelu v prostředí Pythonu, což přináší mnohé výhody, zejména v rychlosti. Další velkou výhodou je, že se jedná o open source produkt. [19]

## 3 ROBOTIC PROCESS AUTOMATION (RPA)

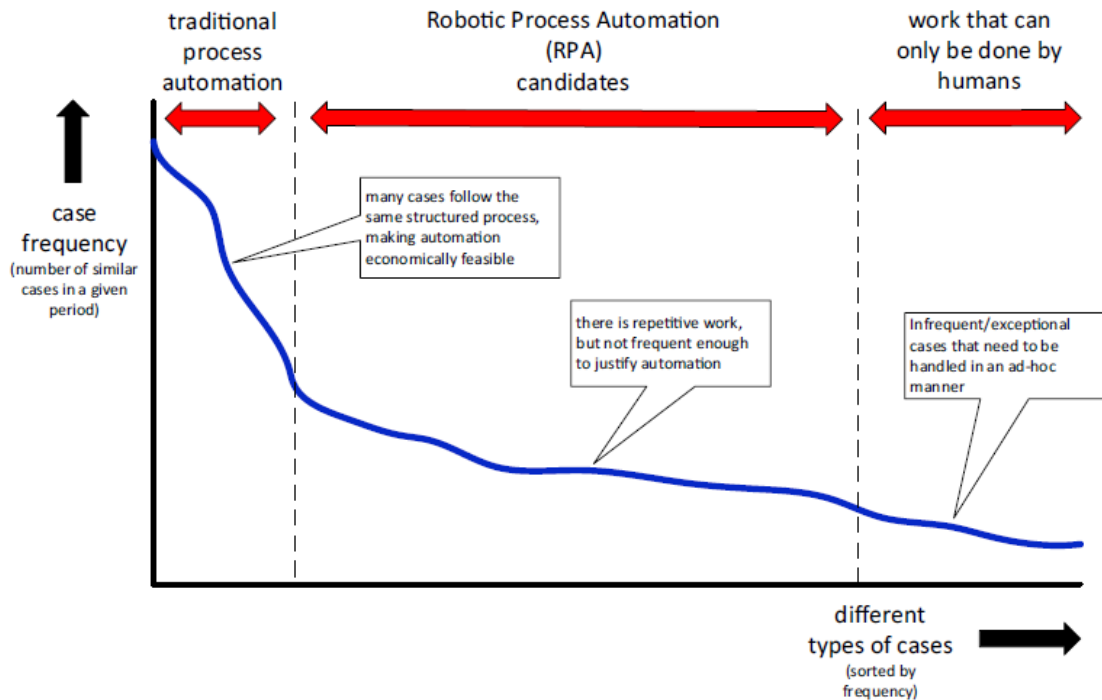
Dalším v poslední době poměrně často skloňovaným tématem je automatizace. Toto téma však není věnováno robotům v továrnách, nýbrž robotům, které automatizují softwarové procesy. Proč je dobré automatizovat tyto procesy? Protože pak jsou ty procesy efektivnější, méně nákladnější, s menší chybovostí a lidská práce, jak tvrdí experti v UI Path, by měla být kreativní a inspirující. [3]

### 3.1 Úvod do automatizace

Jak již bylo zmíněno, pod slovem „robot“ si nepředstavujeme stroje na výrobní lince ve Škoda Auto, ani R2D2 ze Star Wars, nýbrž softwarové řešení podobné této dvojce. Toto řešení má za úkol zautomatizovat proces, který člověk vykonává na PC. Pro automatizaci takového procesu není potřeba programátor, který napíše kód, nýbrž člověk z businessu, znalý daného procesu, který nakonfiguruje robota, aby vykonával tento proces. Zatímco automatizace tradičním způsobem, kdy vývojář napíše kód a ten se posléze implementovat do systému, automatizace pomocí využití RPA ponechá původní systém nezměněný a nahradí člověka v uživatelském rozhraní. Robot pak interaguje s objekty, ať už jsou umístěny na webových stránkách nebo v desktopových aplikacích. [26] [24]

Příkladem může být třeba HR specialista ve velké společnosti, který je vedoucím náboru nových zaměstnanců. Tento proces obvykle vyžaduje logování se do spousty systémů a nastavování profilu zaměstnance pro benefity, payroll, email, firemní kartu a další. Spoustu těchto úkolů by šlo nazvat jako rutinní. Nyní si představme, že bychom měli robota, který tyto všechny rutinní úkoly vykonává, a specialista se věnuje pouze speciálním případům, komunikaci a unikátním procesům, které automatizovat nejdou. Znamenalo by to velkou časovou úsporu, zkvalitnění služeb, a ve finále finanční úsporu, jelikož tento proces bude zvládat s méně lidmi. [26]

Otázkou zůstává: které procesy automatizovat? A jakým způsobem? Odpověď nalezneme na obrázku 26. Pro velmi častý výskyt stejných (nebo velice podobných) případů (procesů) je vhodné zvolit tradiční proces automatizace spojený s implementováním kódu do systému. Pro méně časté a více specifické procesy (či části procesů) je vhodné použít RPA. Lidský faktor nám pak zbývá na málo časté a velmi specifické procesy, které nelze automatizovat, nebo kvůli jejich malému výskytu se jejich automatizace nevyplatí. [24]



**Obrázek 26:** Kdy automatizovat?

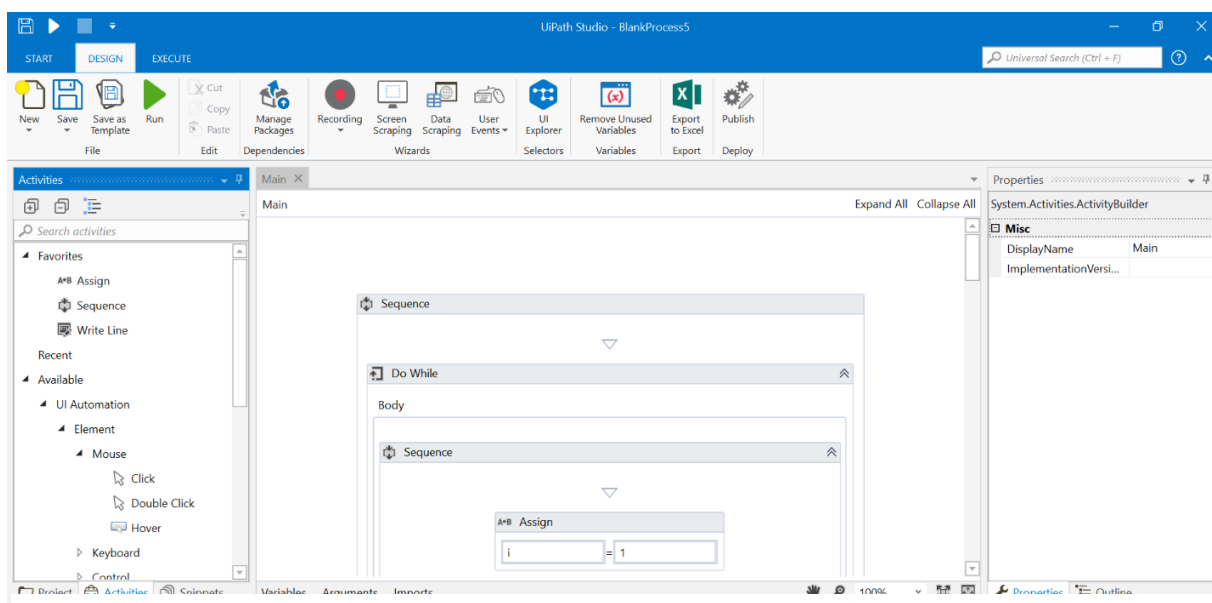
*Zdroj: [24]*

Pokud si představíme společnost se všemi jejími divizemi a procesy, možná nás napadne, kdo je tedy ten, kdo přijde a řekne: „Tento proces by to chtělo automatizovat.“ V tomto kontextu se setkáváme s termínem RPA analyst (RPA analytik). Člověk na této pozici proaktivně zkoumá procesy v rámci organizace a na základě jeho zkušeností identifikuje, které procesy je vhodné automatizovat. Dále se podílí na sepsání dokumentace, obsahující detailní RPA požadavky. Tato pozice se mírně podobá business analytikovi. Rozdílem však je, že business analytik proaktivně nevyhledává procesy k automatizaci, ale předává IT požadavky lidí z businessu, aby byly zaimplementovány do IT ekosystému. [26]

### 3.2 UI Path

UI Path je jedna z platforem, které umožňují tvorbu a konfiguraci robotů. Tento robot pak může interagovat s webovými stránkami či aplikacemi, s desktopovými aplikacemi, či komunikovat s CRM systémy nebo databázemi. Konfigurace takového robota probíhá pomocí sestavení takzvaného workflow, které se skládá z jednotlivých kroků daného procesu. Tento proces je možné nahrát pomocí tlačítka “Recording” a jednotlivé funkce budou i s parametry následně automatiky vygenerovány. Další možností je manuálně tyto funkce vložit a nastavit parametry.

Funkcí, které nám prostředí UI Path nabízí, je opravdu nespočet. Je možné vybrat funkce od klikání na tlačítka, klikání na text, čtení textu, psaní textu až po rozeznání textu pomocí OCR či automatické rozeznání a načítání dat v tabulkové či jiné struktuře z webu. Klíčové je správně identifikovat objekt, s kterým má robot interagovat. V prostředí UI Path toto provádíme pomocí selectoru, který obsahuje html kód daného objektu. Tyto selectory pak můžou být buď statické (pevně daný kód) nebo dynamické (kód se může měnit). Vedle toho UI Path nabízí funkcionality, na které jsme zvyklí s programovacími jazyky, jako práce s proměnnými, cykly či kontejnery. Pro názornost je na obrázku 27 zobrazeno prostředí platformy UI Path.



Obrázek 27: UI Path prostředí

*Zdroj: Vlastní zpracování v softwaru UI Path*

Nutno dodat, že společnost UI Path nám nabízí dvě verze jejich produktu. **Enterprise platforma** je cílena na velké společnosti a hlavním benefitem je podpora cloudu a kolaborace více vývojářů na více robotech. Dalším benefitem je rozšířená podpora ze strany expertů UI Path. Tato verze je samozřejmě zpoplatněna. Vedle toho je nabízena **Community edice**, která je cílená na mikro společnosti, akademiky či jednotlivce, a je zcela zdarma. Na cloud je pak možné připojit pouze 2 roboty a podpora je možná pouze formou community fóra. [3]

## 4 PRAKTICKÁ ČÁST

Praktická část (respektive její jádro), jak již napovídá název diplomové práce, bude věnována modelování v prostředí investic.

### 4.1 Úvod do problému

Jak již bylo zmíněno v zadání diplomové práce, cílem je sestavit model, který určí poměr aktiv, které je vhodné držet v portfoliu. Vzhledem k tomu, že druhů aktiv je nespočet, pojdme se vymežit na jednotlivé sektory americké ekonomiky. Model tedy bude modelovat, jak mít nejvýhodněji rozložené portfolio v rámci sektorů. Aktiva si poté můžeme představit buď ETF fondy, které kopírují vývoj sektoru, nebo nakoupené jednotlivé tituly z indexu S&P 500. Nutno dodat, že v některých situacích je nejvýhodnější nedržet žádné portfolio a mít pouze peníze na účtech.

Akcie jsou zvoleny z amerického trhu, jelikož americký trh je nejefektivnější (pokud bereme v potaz jeho podobu za posledních několik desítek let, nikoliv 20. léta 20. století) a zároveň je k dispozici nejdelší časová řada relevantních dat.

Vezměme v potaz fakt, že aktivně řízené fondy (portfolio manažeři vybírají akcie, případně jiná aktiva, která budou držet v portfoliu) ve velké většině případů zaostávají za svým benchmarkem (indexem). Vzhledem k tomu, že nechceme být touto většinou a nejsme ani Warren Buffet, který má “na esemesce” CEO Coca-Cola a má skvělou ruku na výběr akcií, pak strategie bude investovat do indexu, konkrétně ETF fondů. Podíl akcií v portfoliu pak bude určen rizikovostí držby akcií, z důvodu překoupenosti trhu či nadhodnocení akcií.

Toto riziko pak budeme modelovat ze dvou pohledů: míry agregace akciových titulů a investičním horizontu. V rámci agregace akciových titulů se budeme dívat na trh jako celek, jednotlivé sektory, až se dostaneme na úroveň jednotlivých akciových titulů (obsažených v indexu S&P 500 k 30.3.2019). V rámci investičního horizontu budeme mít krátkodobý pohled, tedy jestli akcie v horizontu jednoho měsíce budou ztrátové, a dlouhodobý pohled, tedy jestli budou ztrátové v horizontu jednoho roku.

### 4.2 Vlastní přístup

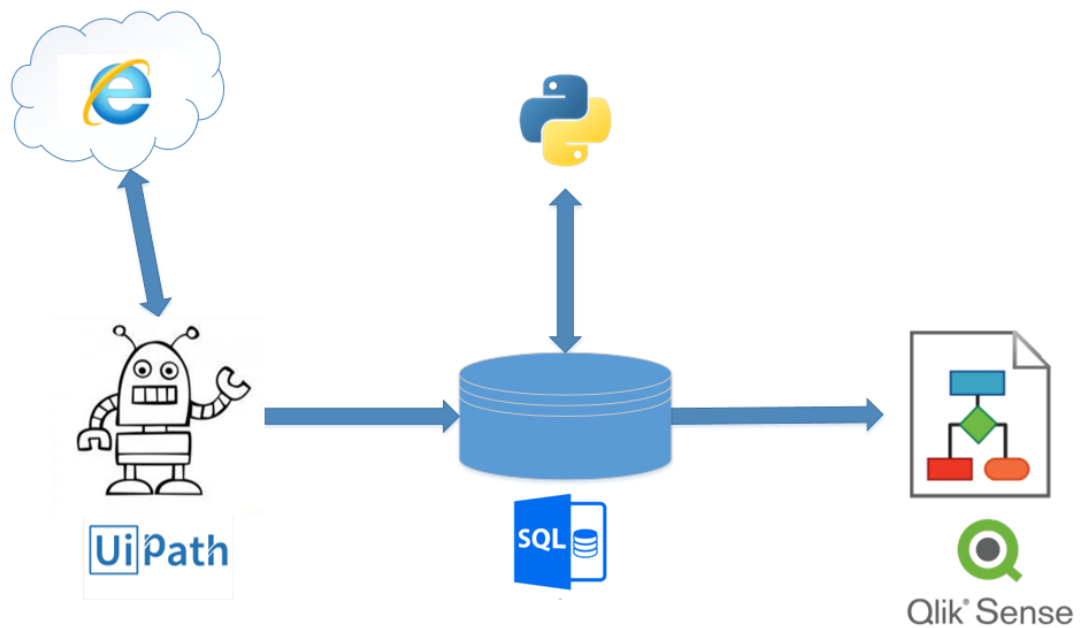
Vzhledem k tomu, že časová řada, s kterou budeme operovat je poměrně dlouhá, firem, za které budeme analyzovat tržní data je poměrně mnoho a dat máme poměrně širokou škálu,



dostáváme se na miliony záznamů a klasický Excel není zrovna ideální na práci s takovýmto objemem dat.

Proto jako uložště dat byla zvolena SQL databáze a analýza bude probíhat v prostředí programovacího jazyka Python za pomoci modulů Numpy, Pandas a SciPy. Výstupy poté budou pro lepší přehlednost vizualizovány do dashboardu pomocí Business Intelligence nástroje Qlik Sense.

Jelikož model bude využívat velké množství dat a ty pro aktuálnost modelu budou potřeba být aktualizovány na denní bázi, bylo by velice časově náročné toto dělat ručně. Pro tento účel budou nakonfigurováni roboti v prostředí UI Path, kteří potřebná data vyhledají, stáhnou a načtou do modelu „na jedno tlačítko“. Celý ekosystém bude graficky vypadat jako na obrázku číslo 28.



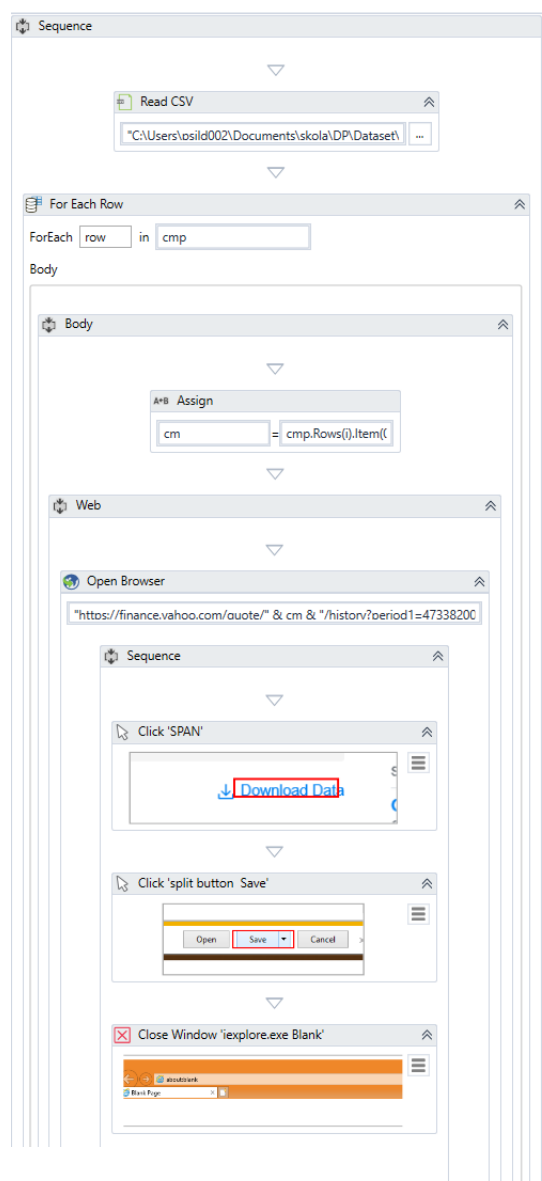
**Obrázek 28:** Vlastní přístup

*Zdroj: Vlastní zpracování v softwaru Visio Professional*

## 4.2.1 Sběr dat (Data collection)

### Tržní data

V první řadě byl potřeba získat seznam tickerů (označení na burze) 505 firem, ze kterých se skládá index S&P 500, a tento seznam uložit do csv souboru. Robot si poté tento soubor načte a po jedné firmě stahuje data. Zdrojem je web Yahoo [27] a na firmu se odkazuje právě pomocí tickeru (proměnná `cm`) vložením do URL adresy: "[https://finance.yahoo.com/quote/"&cm&"/history?period1=473382000&period2=1554933600&interval=1d&filter=history&frequency=1d](https://finance.yahoo.com/quote/)". Algoritmus pak v prostředí UI Path vypadá takto:



Obrázek 29: Tržní data v UI Path

Zdroj: vlastní zpracování v prostředí softwaru UI Path

## *Makroekonomická data*

Obdobným algoritmem, jako jsme získali tržní data, jsou získány také makroekonomická data. Výhodou je, že není potřeba 505krát opakovat cyklus načtení a stažení dat, nýbrž pro každý indikátor pouze jednou. Konkrétně jde o měnový agregát **M2** [5], **Index spotřebitelských očekávání** [12] a úroková míra určená pomocí **Výnosových křivek** [23].

## *Ostatní data*

Opět obdobným algoritmem jsou získány další data, které jsou vhodné k modelování. Patří mezi ně **Index S&P 500** [27], Vývoj **podnikatelských a spotřebitelských úvěrů** [4], a vývoj **indexu zlata** [7].

Nakonec jsou potřeba data ohledně zařazení firem do **sektorů** [27] a jejich **procentuální podíl** [17] na indexu S&P 500. Toto bylo provedeno jednorázovým extrahováním dat z tabulky a následném uložení do csv souboru.

### **4.2.2 Skladování dat (Data warehouse)**

Dalším krokem je nahrání dat do databáze. Tento krok byl proveden v prostředí Microsoft SQL Server Management Studio pomocí následujících dotazů:

```
CREATE TABLE YC(  
Date DATE,  
1YEAR FLOAT,  
2YEAR FLOAT,  
3YEAR FLOAT,  
5YEAR FLOAT,  
7YEAR FLOAT,  
10YEAR FLOAT,  
20YEAR FLOAT,  
30YEAR FLOAT  
)
```

Tyto kroky slouží k vytvoření tabulky k nahrání dat. Zde je uvedena ukázka pro data za výnosové křivky. Tady je klíčové správně nadefinovat typy proměnných, abychom neukládali data, či číselné hodnoty do sloupce nadefinovaného jako text.

```
BULK INSERT YC  
FROM 'C:\Users\sildp\Downloads\FED-SVENY.csv'  
WITH  
(  
    FIRSTROW = 2,  
    FIELDTERMINATOR = ',', --CSV field delimiter  
    ROWTERMINATOR = '\n', --Use to shift the control to next row  
    TABLOCK  
)
```

Výstup následně nahraje data z csv souboru do dané tabulky.

Vzhledem k tomu, že není moc praktické nahrát 505 souborů do databáze a pak mít tržní data v 505 tabulkách, je třeba před nahráním do databáze tyto soubory spojit a identifikovat je dle nově přidané proměnné vyjadřující společnost, za kterou jsou data („company“). Pro tento účel byl využit opět jazyk Python:

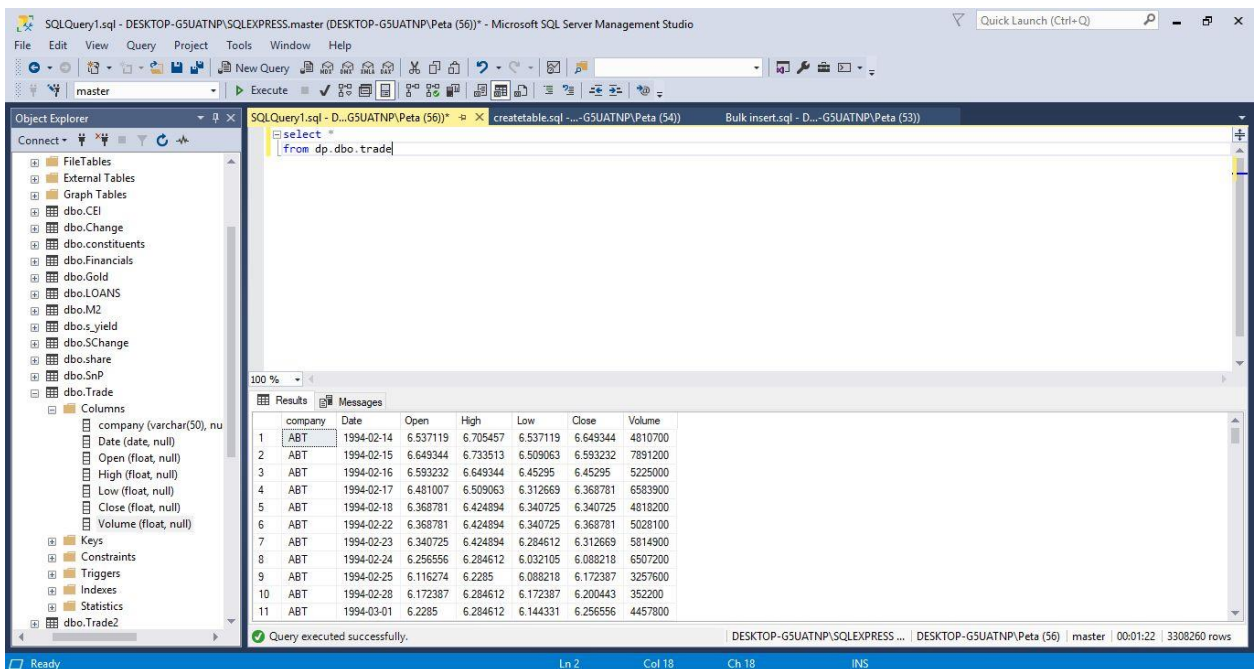
```
import numpy as np
import pandas as pd
import glob

df = pd.read_csv('Trade\A.csv', delimiter=',')
i=0
df.insert(0,"company","A",1)
df.to_csv('Trade.csv')

all_files = glob.glob("Trade/*.csv")

for filename in all_files:
    df = pd.read_csv(filename, delimiter=',')
    i+=1
    print(round(i/505*100,2))
    name = filename.replace("Trade\\", "").replace(".csv", "")
    if name != "A":
        df.insert(0,"company", name,1)
        df.to_csv('Trade.csv', mode = 'a', header=False)
```

Dataset v prostředí Microsoft SQL Server Management Studio pak bude vypadat takto:



**Obrázek 30:** Dataset v SQL

*Zdroj: Vlastní zpracování v softwaru Microsoft SQL Server Management Studio*

### 4.2.3 Příprava dat (Data preparation)

Vzhledem k tomu, že data jsou stažená již víceméně vyčištěná, není nutné, aby prošla dalším intenzivním čištěním. Co je avšak nezbytné vyřešit, co udělat s chybějícími daty, jelikož index spotřebitelských očekávání a vývoj úvěrů je reportován na měsíční bázi, oproti tržním datům, výnosovým křivkám a měnovým agregátům, které je možné získat na denní bázi.

Jelikož se jedná o okamžikovou časovou řadu, pak chybějící hodnoty dává smysl nahradit pomocí interpolace. Tu, před tím, než začneme modelovat, provedeme pomocí funkce `interpolate()` obsažené v modulu `Pandas`.

V rámci přípravy dat je nutné dodat, že při modelování budoucího vývoje, je třeba počítat s časovým zpožděním se kterým data získáváme. V případě tržních to je pouze jeden den, nicméně v případě indexu spotřebitelských očekávání to mohou být až dva měsíce.

### 4.2.4 Návrh modelu

Jak již bylo zmíněno, data budeme modelovat na 3 úrovních: akcie, sektoru a trhu. Začneme tedy od nejnižší úrovně, která bude také nejméně přesná, a postupně s využitím výsledků modelu půjdeme výše.

#### *Model jednotlivých akcií*

Vzhledem tomu, že vazba makroekonomických faktorů přímo na jednotlivou akci, je poměrně slabá, nedává smysl do modelu pro jednotlivé akcie tyto indikátory zahrnovat. Máme tu ale metody technické analýzy, které můžeme pro tento účel využít. Co tedy budeme modelovat?

Budeme se snažit predikovat výhodnost držení akcie pomocí kupních a prodejních signálů technické analýzy. Jak tedy vydefinujeme výhodnost držení akcie? Vzhledem k tomu, že metody technické analýzy se zaměřují zpravidla na krátké období, budeme modelovat, jestli při držení akcie, budeme o měsíc později realizovat výnos. Adepty na model, který nám bude modelovat proměnnou na intervalu  $\langle 0;1 \rangle$ , máme dva: logistickou regresi a klasifikační/rozhodovací stromy.

#### *Model sektorů*

Do vývoje cen akcií celého odvětví se již vliv makroekonomických indikátorů promítnout může, proto dává smysl ho sem zahrnout. Vedle nich využijeme poznatky z technické analýzy předchozího modelu a jeho výstup zahrneme mezi vstupní proměnné. Makroekonomické

indikátory nasvědčují dlouhodobějšímu charakteru, proto tedy budeme modelovat, jestli při jednoletém držení akcie, budeme realizovat výnos. Opět zde dává smysl využít buď klasifikační/rozhodovací stromy nebo logistickou regresi.

#### *Model celého akciového trhu*

Tento model bude založen na bázi simulace Monte Carlo, kdy chceme z pravděpodobnostního vektoru (který vyjadřuje, jak je výhodné akcie držet) získat poměr aktiv v portfoliu. Budou tedy modelovány různé způsoby přetransformování tohoto vektoru a poté sledována závislost scénářů na budoucím zisku.

#### **4.2.5 Příprava modelu**

Nejdříve začneme s přípravou modelu držby jednotlivých akcií za využití technické analýzy.

#### *Model jednotlivých akcií*

Pro tento účel bylo zvoleno 5 metod technické analýzy: Bollingerovy pásy, Relative strength index, MACD, Stochastik a Money flow index.

V první fázi je potřeba vytvořit pomocné indikátory, jako klouzavé průměry, exponenciální klouzavé průměry a jiné, a poté identifikovat nákupní a prodejní signály pro celou časovou řadu pro každou společnost. Poté určit prodejní a nákupní signály a dle nich rozhodnout, kdy má být daný akciový titul držen a kdy ne. Výsledkem bude 5 (respektive 6 pokud vytvoříme počet metod, dle kterým bychom měli akcii držet) proměnných, nabývajících hodnoty 0 nebo 1.

Dále potřebujeme vytvořit modelovanou proměnnou. Ta bude opět nabývat hodnot 0 a 1, podle toho, jestli se akcii vyplatí nebo nevyplatí držet. Aby tato proměnná často nekolísala a neznamenaloby to velkou spoustu nákupních či prodejních signálů, stanovme hranice, nad které musí hodnota stoupnout, aby se jednalo o nákupní signál, respektive pod které musí klesnout, aby se jednalo o prodejní signál. Tyto hranice jsou expertně stanoveny jako  $\pm 5\%$ .

Modelovanou proměnnou poté dostaneme pomocí jednadvacetidenního klouzavého produktu (součinu) denních změn. V případě, že akcii nedržíme a hodnota produktu stoupne nad hranici 1,05, jedná se o nákupní signál a proměnná nabývá hodnoty 1. Tuto hodnotu má až do doby prodejního signálu, tedy když hodnota klouzavého produktu klesne pod hranici 0,95.

Na základě těchto proměnných vytvoříme model. V prostředí Pythonu to můžeme udělat pomocí následujícího kódu:

```
import pyodbc
import numpy as np
import pandas as pd
from sklearn import metrics
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import classification_report
from sklearn import tree
import graphviz
from sklearn.tree import export_graphviz
from sklearn.externals.six import StringIO
from sklearn.model_selection import train_test_split

#Functions for identification holding share
def bol(row, i1, i2, i3, out):
    if i1[row] <= i2[row] and i1[row-1] > i2[row-1]:
        val = 0
    elif i1[row] >= i3[row] and i1[row-1] < i3[row-1]:
        val = 1
    else:
        val = out[row-1]
    return val

def rs(row, i, out, c):
    if i[row] <= c:
        val = 1
    elif i[row] >= 100-c:
        val = 0
    else:
        val = out[row-1]
    return val

#Database connection
cnxn = pyodbc.connect("Driver={SQL Server Native Client 11.0};"
                    "Server=DESKTOP-G5UATNP\\SQLEXPRESS;"
                    "Database=DP;"
                    "Trusted_Connection=yes;")

query = ("select distinct [company] from dp.dbo.trade2 order by company")
companies = np.array(pd.read_sql_query(query,cnxn))

query = ("select [date], [Close], [High], [Low], [Volume], [Open], [company] from
dp.dbo.trade order by company,date")
all_data = pd.read_sql_query(query,cnxn)
print('Loaded')

#Calculating indicators for each company
for comp in range(0,len(companies)):
    c = companies[comp,0]
    trade_data = all_data[all_data['company'] == c]
    trade_data=trade_data.reset_index()
    CLOSE = np.array(trade_data['Close'])
    Change = pd.DataFrame(trade_data['Close'] / trade_data['Open']-1)
    Change2 = pd.DataFrame((Change+1).apply(np.floor))
    RSN = np.array(Change2.rolling(window=20).sum())
    RSI = np.array(100-(100/(1+RSN/(20-RSN))))
    MA = np.array(trade_data['Close'].rolling(window=20).mean())
    EMA = pd.DataFrame(trade_data['Close'].ewm(span=26, adjust=False).mean() -
trade_data['Close'].ewm(span=12, adjust=False).mean())
```

```

TRIGGER = np.array(EMA.ewm(span=9, adjust=False).mean())
EMA = np.array(EMA)
STD = np.array(trade_data['Close'].rolling(window=20).std())
STOCHASTIC = pd.DataFrame(((trade_data['Close'] -
trade_data['Low'].rolling(window=14).min())/(trade_data['High'].rolling(window=14).max
()-trade_data['Low'].rolling(window=14).min())*100))
S_TRIGGER = np.array(STOCHASTIC.rolling(window=3).mean())
STOCHASTIC = np.array(STOCHASTIC)
MRH = np.where(Change2[0]==1,trade_data['Volume'],0)
MRL = np.where(np.array(Change2[0])==0,trade_data['Volume'],0)
MFI = np.array(100-100/(1+(pd.DataFrame(MRH).rolling(window=14).sum() /
pd.DataFrame(MRL).rolling(window=14).sum()))))
BU = MA + 2*STD
BD = MA - 2*STD

#Preparing arrays for technical indicators
HOLD_BOL = np.ones(len(trade_data))
HOLD_RSI = np.ones(len(trade_data))
HOLD_MACD = np.ones(len(trade_data))
HOLD_STOCH = np.ones(len(trade_data))
HOLD_MFI = np.ones(len(trade_data))
HOLD_SUM = np.zeros(len(trade_data))

#Preparing target value (Yield)
Yield = (Change+1).rolling(window=w).apply(np.prod)
Yield = np.array(Yield)
l=len(Yield)
Yield21=np.ones(l)
for i in range (20,l):
    if (Yield[i] > 1.05 and Yield21[i-1] == 0) or (Yield[i] > 0.95 and Yield21[i-
1] == 1):
        Yield21[i] = 1
    else:
        Yield21[i] = 0
l=l-w

#Applying function to decide when to hold
for i in range(19,len(HOLD_BOL)):
    HOLD_BOL[i] = bol(i, CLOSE, BD, BU, HOLD_BOL)
    HOLD_RSI[i] = rs(i,RSI,HOLD_RSI,30)
    HOLD_MACD[i] = bol(i, EMA, TRIGGER, TRIGGER, HOLD_MACD)
    HOLD_STOCH[i] = bol(i, STOCHASTIC, S_TRIGGER, S_TRIGGER, HOLD_STOCH)
    HOLD_MFI[i] = rs(i,MFI,HOLD_MFI,20)
HOLD_SUM = HOLD_BOL + HOLD_RSI + HOLD_MACD + HOLD_STOCH + HOLD_MFI

#x matrix definition
x = pd.DataFrame(data = (HOLD_BOL, HOLD_RSI, HOLD_MACD, HOLD_STOCH, HOLD_MFI,
HOLD_SUM)).T
X_train, X_test, y_train, y_test = train_test_split(
    x.iloc[19:l,:], Yield21[19+w:], test_size=0, random_state=42)

#Logistic regression implementation
logr = LogisticRegression(random_state=0, solver='lbfgs',
                           multi_class='multinomial').fit(X_train, y_train)
predictions = logr.predict(X_train)
conf = logr.predict_proba(x.iloc[19:l,:])
acc[comp]=metrics.accuracy_score(y_train, predictions)
#print(classification_report(y_train,predictions))
#print(acc)
#print(conf)
#print(predictions)

```



```

#exporting
export = pd.DataFrame(data = (trade_data.iloc[19:,0], trade_data.iloc[19:,1],
trade_data.iloc[19:,5], HOLD_BOL, HOLD_RSI, HOLD_MACD, HOLD_STOCH, HOLD_MFI, HOLD_SUM,
np.append(np.zeros(19),conf[:,1])))
export['company'] = c
if comp > 0:
    export.to_csv('exp.csv')
else:
    export.to_csv('exp.csv', mode='a', header=False)
print(round(comp/7*100,2))
del export
np.savetxt('prob5.csv',acc.reshape(-1,1), delimiter = ',')

```

Výsledkem je model logistické regrese, který na základě indikátorů technické analýzy modeluje pravděpodobnost, jestli každou z 505 akcií z indexu S&P 500, je výhodné držet.

### *Model sektorů*

Zde je potřeba nejdříve získat modelovanou proměnnou, tedy výnos dle sektoru. Potřebujeme tedy vážený průměr výnosů firem jejich váhami zastoupení v indexu S&P 500 v rámci každého sektoru. V Pythonu je to možné pomocí následujícího kódu:

```

import pyodbc
import numpy as np
import pandas as pd

#Database connection
cnxn = pyodbc.connect("Driver={SQL Server Native Client 11.0};"
                    "Server=DESKTOP-G5UATNP\SQLEXPRESS;"
                    "Database=DP;"
                    "Trusted_Connection=yes;")

i=0

#Data load
query = ("select Symbol, share, Sector from constituents")
df = pd.read_sql_query(query,cnxn)
companies = np.array(df['Symbol'])
df = df.set_index('Symbol')
shares = df.to_dict()

query = ("select distinct [Date] from dp.dbo.trade as t order by Date")
df = pd.read_sql_query(query,cnxn)
dates = np.array(df['Date'])

query = ("select distinct Sector from constituents")
df = pd.read_sql_query(query,cnxn)
sectors = list(df['Sector'])

query = ("select [Date], [company], [Unnamed_6] from dp.dbo.exp order by [Date]")
trade_data = pd.read_sql_query(query,cnxn)

c=0

#Array preparing
weights = pd.DataFrame(index = dates, columns = ['Total'], dtype='float')
weights = weights.fillna(0)
t_change = pd.DataFrame(index = dates, columns = ['Total'], dtype='float')
t_change = t_change.fillna(0)

```

```

#Array fill by day
for i in range(0,len(dates)):
    day = np.array(trade_data[trade_data['Date'] == dates[i]])
    for j in range(0,len(day)):
        comp = day[j,1]
        weight = shares['share'][comp]

        if weights['Total'][dates[i]]==0:
            weights['Total'][dates[i]] = weight
        else:
            weights['Total'][dates[i]] = weights['Total'][dates[i]] + weight

        if t_change['Total'][dates[i]]==0:
            t_change['Total'][dates[i]] = (weight * (day[j,2]))
        else:
            t_change['Total'][dates[i]] = t_change['Total'][dates[i]] + (weight *
(day[j,2]))
            print(i/len(dates)*100)
chng = pd.DataFrame(t_change/weights)
chng.to_csv('s_change.csv')

```

V další fázi je třeba načíst proměnné a implementovat model:

```

import pyodbc
import numpy as np
import pandas as pd
import sklearn
from sklearn import metrics
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import classification_report
from sklearn import tree
import graphviz
from sklearn.tree import export_graphviz
from sklearn.externals.six import StringIO

#Database connection
cnxn = pyodbc.connect("Driver={SQL Server Native Client 11.0};"
                    "Server=DESKTOP-G5UATNP\SQLEXPRESS;"
                    "Database=DP;"
                    "Trusted_Connection=yes;")

#Data load
query = ("select * from dp.dbo.s_prob where Date > '1989-12-31' and Date < '2018-04-13' order by Date")
df = pd.read_sql_query(query,cnxn)
TA = df.melt(id_vars = ['Date'], var_name = ['sector'])

query = ("select * from dp.dbo.SnP as s full join DP.dbo.M2 as m on s.Date = m.Date
full join dp.dbo.YC as y on s.Date = y.Date full join dp.dbo.Gold as g on s.Date =
g.Date full join dp.dbo.CEI as c on s.Date = c.Date2 full join dp.dbo.LOANS as l on
s.Date = l.Date2 where s.Date > '1989-12-31' and s.Date < '2019-04-13' and (LOCATION =
'USA' or LOCATION is null) order by m.Date")
df = pd.read_sql_query(query,cnxn)

mp = np.array(df['M2'])
gold = np.array(df['Gold']/df['M2'])
yc = np.array(df['Index'].interpolate())
cl = np.array(df['Close'])
cei = np.array(df['CEI'].interpolate())
loans = np.array(df['Loans'].interpolate()/df['M2'])
x = pd.DataFrame(data= (mp,yc,gold,cei,loans)).T

```

```

query = ("select distinct sector from dp.dbo.s_yield")
df = pd.read_sql_query(query,cnxn)
sectors = np.array(df['sector'])
pred = pd.DataFrame()

#Model implementation
for s in sectors:
    STA = TA[TA['sector'] == s]['value']
    STA = STA.reset_index()
    x['TA'] = STA['value']

    query = ("select y21, [Index] from DP.dbo.s_yield where Sector = '" + s + "' and
Date < '2019-04-13' and Date > '1989-12-31' order by Date")
    df = pd.read_sql_query(query,cnxn)
    hold = np.array(df['y21']+1)
    for i in range (0,len(hold)):
        if (hold[i] > 1.05 and hold[i-1] == 0) or (hold[i] > 0.95 and hold[i-1] == 1):
            hold[i] = 1
        else:
            hold[i] = 0
    x[0] = df['Index']/mp

# x_train, x_test, y_train, y_test = train_test_split(x, hold, test_size=0.33)

clf = tree.DecisionTreeClassifier(criterion="gini", max_depth=5)
clf = clf.fit(x, hold)
predictions = pd.DataFrame(clf.predict_proba(x))
pred[s] = predictions[25]
#print(classification_report(hold,predictions))
#print("Accuracy:",metrics.accuracy_score(hold, predictions))
dot_data = StringIO()
dotfile = open(s + ".dot", 'w')
tree.export_graphviz(clf, out_file=dotfile)
dotfile.close()
pred['date'] = np.array(TA2.loc[TA2['sector'] == 'Total','Date'])
pred = pred.set_index('date')
pred.to_csv('dp.csv')

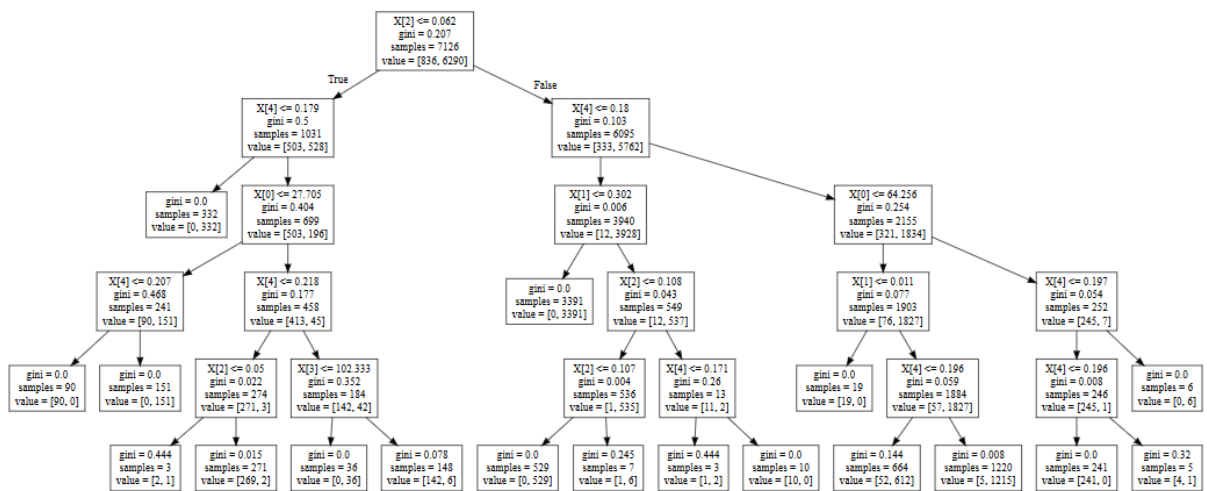
```

Pro tento případ byl zvolen model klasifikačních stromů. Jako kritérium dělení je zvolen gini koeficient. Po testování, aby strom nebyl „overfitting“ byla zvolena maximální hloubka 5. Na obrázku 31 je ukázka klasifikačního reportu a dále na obrázku 32 rozhodovacího stromu pro sektor Informačních technologií.

	precision	recall	f1-score	support
0.0	0.99	0.93	0.96	836
1.0	0.99	1.00	0.99	6290
micro avg	0.99	0.99	0.99	7126
macro avg	0.99	0.96	0.98	7126
weighted avg	0.99	0.99	0.99	7126
Accuracy: 0.9903171484703901				

**Obrázek 31:** Klasifikační report IT

*Zdroj: Vlastní zpracování v prostředí Microsoft Visual Studio*



**Obrázek 32:** Klasifikační strom IT

*Zdroj: Vlastní zpracování v prostředí Microsoft Visual Studio*

Výstupem z tohoto modelu je pravděpodobnostní vektor, který určuje pravděpodobnost, že při nabytí daných hodnot, se vyplatí držet index daného sektoru.

### Model akciového trhu

Tento model vychází z výsledků předchozího modelu, který modeluje pravděpodobnost výhodnosti držení akcií daného sektoru po dnech. Tyto pravděpodobnosti však potřebujeme přetransformovat v podíly portfolia. Vezmeme tedy pro každý den pravděpodobnostní vektor z minulého modelu a každou jeho hodnotu vydělíme součtem hodnot tohoto vektoru. Získáme tím vektor vah, jehož suma je rovna 1.

Pojďme ale tuto pravděpodobnost před transformací ještě upravit o konstantu z intervalu  $(-1;1)$ . Když se konstanta limitně bude blížit 1, náš model se bude blížit naivní diverzifikaci (stejný podíl ve všech aktivech). Čím více se bude blížit -1, tím více bude model penalizovat sektory s nízkou pravděpodobností výnosu ve prospěch sektorů s vyšší pravděpodobností. Nutno dodat, že pravděpodobnost nemůže klesnout do záporných hodnot (respektive minimálně na „technickou nulu“) a ani překročit hodnotu 1.

Dále ještě by bylo vhodné ošetřit případy, kdy nám model určil u sektorů nízké pravděpodobnosti, jelikož držet jakýkoliv sektor by bylo rizikové. Po implementaci předchozího bodu by vznikla naivní diverzifikace a portfolio by klesalo. Definujme tedy hranici na intervalu  $(0;1)$ , pod kterou když spadne průměrná pravděpodobnost za všechny sektory v daný den, investor celé své portfolio prodá.

Otázkou však zůstává, jak tyto hodnoty určit? V této otázce nám pomůže metoda Monte Carlo, kdy vygenerujeme velké množství scénářů s náhodnými hodnotami výše zmíněných proměnných na daných intervalech. Poté budeme pozorovat vztah hodnot těchto proměnných a výnosnosti portfolia. Metodu Monte Carlo budeme aplikovat opět v prostředí Pythonu pomocí následujícího kódu:

```
import pyodbc
import numpy as np
import pandas as pd
from random import random

#Database connection
cnxn = pyodbc.connect("Driver={SQL Server Native Client 11.0};"
                    "Server=DESKTOP-G5UATNP\SQLEXPRESS;"
                    "Database=DP;"
                    "Trusted_Connection=yes;")

#Probabilities load
query = ("select [Date],[Consumer Discretionary],[Consumer Staples],[Information
Technology],[Utilities],[Materials],[Financials],[Telecommunication Services],[Real
Estate],[Energy],[Health Care],[Industrials] FROM [DP].[dbo].[sector_probabilities]
where Date < '2019-03-13' and Date > '1989-12-31' order by Date")
df = pd.read_sql_query(query,cnxn)
p_data = df.set_index('Date')
p_data = np.array(p_data)

#Yield by sector and day load
query = ("select [Date],[Consumer Discretionary],[Consumer Staples],[Information
Technology],[Utilities],[Materials],[Financials],[Telecommunication Services],[Real
Estate],[Energy],[Health Care],[Industrials] FROM [DP].[dbo].[SChange] where Date <
'2019-03-13' and Date > '1989-12-31' order by Date")
df = pd.read_sql_query(query,cnxn)
df = df.set_index('Date')
y = np.array(df)

#Variables setting
scenarios=200
yd = np.zeros(len(y))
exp = pd.DataFrame(columns = ['c', 'intercept', 'y'], index=range(0,scenarios))

#Generating scenarios
for i in range(0,scenarios):
    #generating random numbers
    c = random()*2-1
    intercept = random()

    #implementing constant
    data = p_data+c
    data[data < 0] = 0.0001
    data[data > 1] = 1
    data = np.array(data)
    j=0

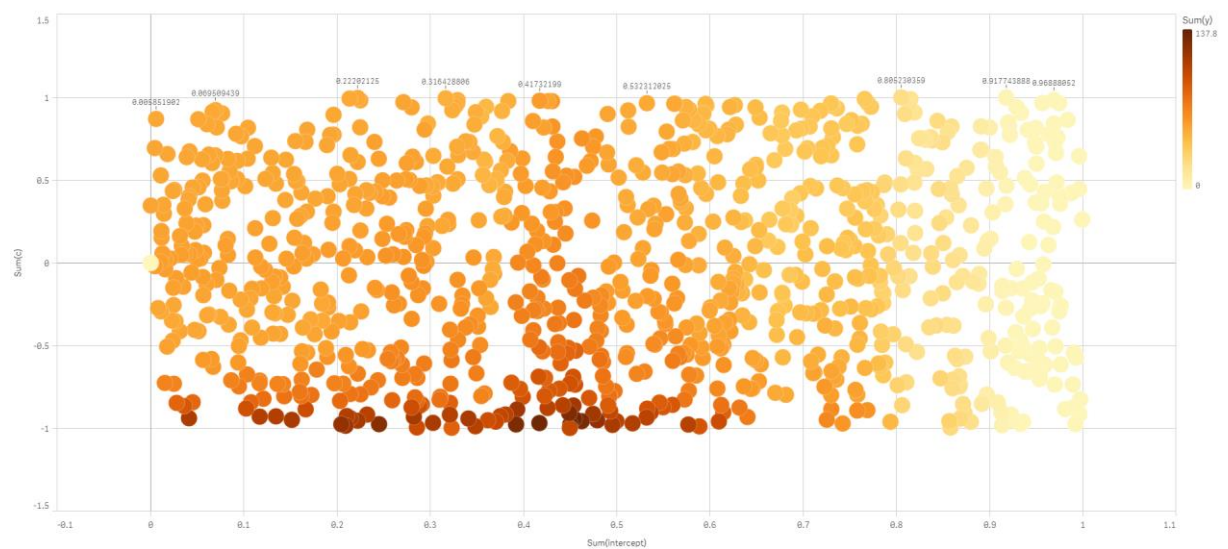
    #walking through rows
    for r in data:
        #implementing intercept
```

```

if np.average(p_data[j,:]) < intercept:
    yd[18] = 0
else:
    #calculating weights and weighted yield
    r = r/r.sum()
    r = r*y[j,:]
    yd[18] = r.sum()
    j+=1
yd = yd+1
#exporting scenario outputs to DataFrame
exp.iloc[i,0] = c
exp.iloc[i,1] = intercept
exp.iloc[i,2] = yd.prod()
if i%1 == 0:
    print(i/scenarios*100)
#exporting DataFrame
pd.DataFrame(exp).to_csv('p2.csv')

```

Výslednou závislost si vizualizujeme pomocí softwaru Qlik Sense:



**Obrázek 33:** Qlik Sense: závislost konstanty a hranice na výnosu

*Zdroj: Vlastní zpracování v softwaru Qlik Sense*

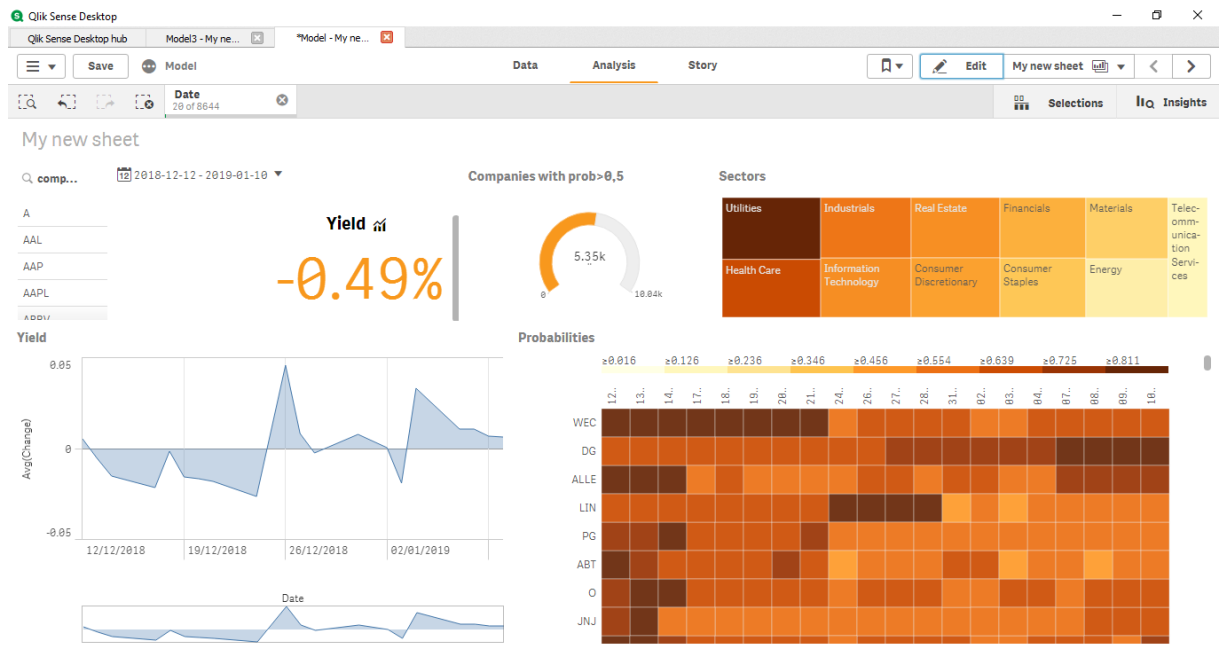
Na grafu věnujeme pozornost výskytu tmavých bodů, který jak můžeme vidět není zcela náhodný. Tyto hodnoty se nejvíce vyskytují kolem hranice 0,45 a konstanty blíží se -1.

Máme tedy 3 modely, které jsou na sobě navázané a na jejichž konci stojí složení portfolia investora. Investor může dále analyzovat jednotlivá rizika v rámci sektorů, trhu či přímo na úrovni jednotlivých akcií. Na základě těchto dat investor může sestavit portfolio vlastní.

## 4.2.6 Vizualizace

Poslední fází je vizualizace poznatků získaných z modelu. Vizualizace provedeme v prostředí Qlik Sense přípravou takzvaných dashboardů. Cílem vizualizace je na první pohled získat co nejvíce informací z dat.

### Akciový model



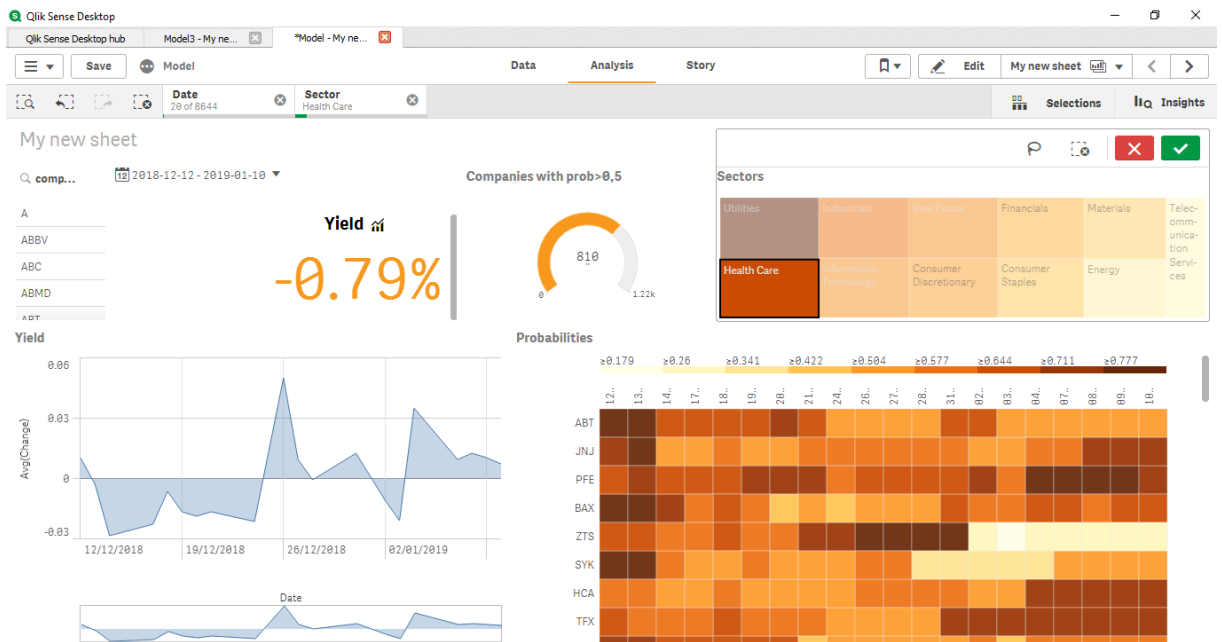
Obrázek 34: Vizualizace: akciový model

*Zdroj: Vlastní zpracování v programu Qlik Sense*

Na obrázku 34 je možné vidět vizualizaci akciového modelu. Vpravo dole vidíme heatmapu, kde je odstínem barvy zobrazena pravděpodobnost z modelu za vybrané období po jednotlivých akciových titulech, seřazené sestupně. V těchto titulech lze buď manuálně listovat nebo filtrovat pomocí filtru nahoře vlevo. Další důležitý filtr je hned vedle něj, který slouží pro výběr období.

Vlevo dole můžeme vidět vývoj průměrných denních změn za všechny akciové tituly. Vedle filtrů můžeme nalézt výnos či ztrátu za zvolené období a napravo od tohoto ukazatele vidíme poměr mezi společnostmi, u kterých je pravděpodobnost z modelu vyšší než 0,5 a u kterých nižší.

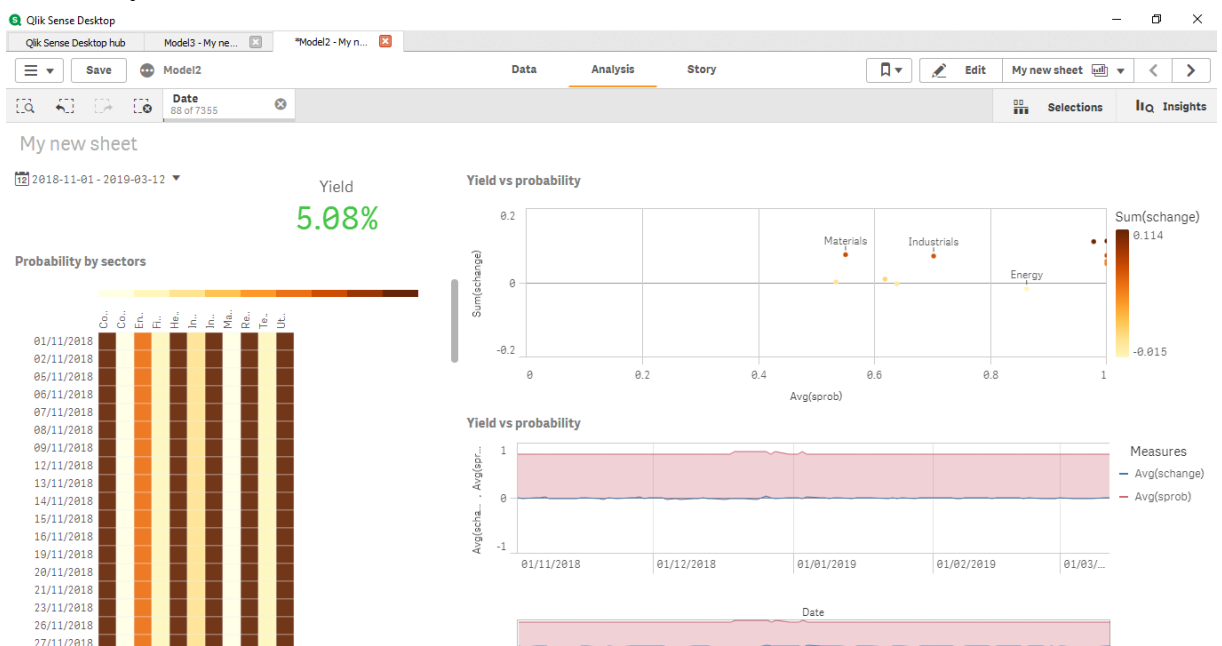
Posledním ukazatelem je rozdělení sektorů, kde velikost a odstín políčka značí průměrnou pravděpodobnost. Výhodou je, že téměř úplně každý objekt umí po kliknutí filtrovat. Například pokud si rozklikneme sektor zdravotní péče (obrázek 35), zobrazí se nám data pouze za společnosti ze sektoru zdravotní péče.



**Obrázek 35:** Vizualizace: akciový model – zdravotní péče

*Zdroj: Vlastní zpracování v programu Qlik Sense*

## Sektorový model



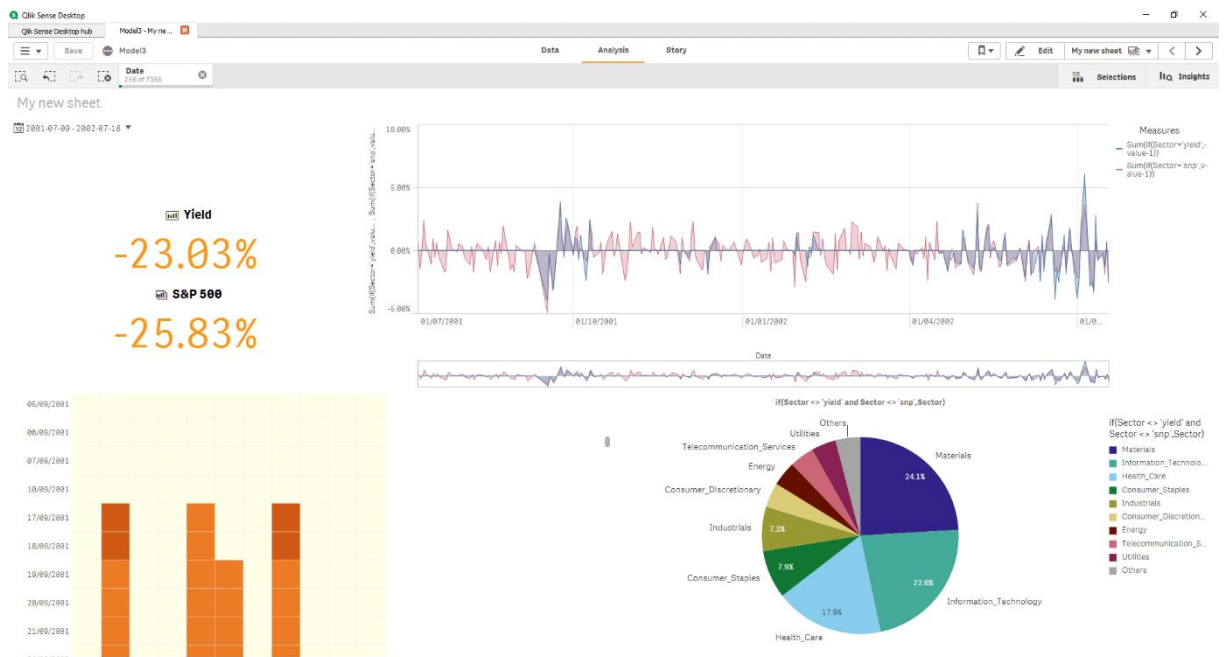
**Obrázek 36:** Vizualizace: sektorový model

*Zdroj: Vlastní zpracování v programu Qlik Sense*

Obrázek 36 obsahuje výstupy ze sektorového modelu. Vlevo dole je vidět opět heatmapa, akorát zde akciové tituly nahradily sektory. Opět zde máme výnos a filtr období. Novým prvkem je zde bodový grafy vyjadřující vztah mezi pravděpodobností, že se vyplatí držet akcie a reálným výnosem.



## Model celého trhu



Obrázek 37: Vizualizace: celý trh

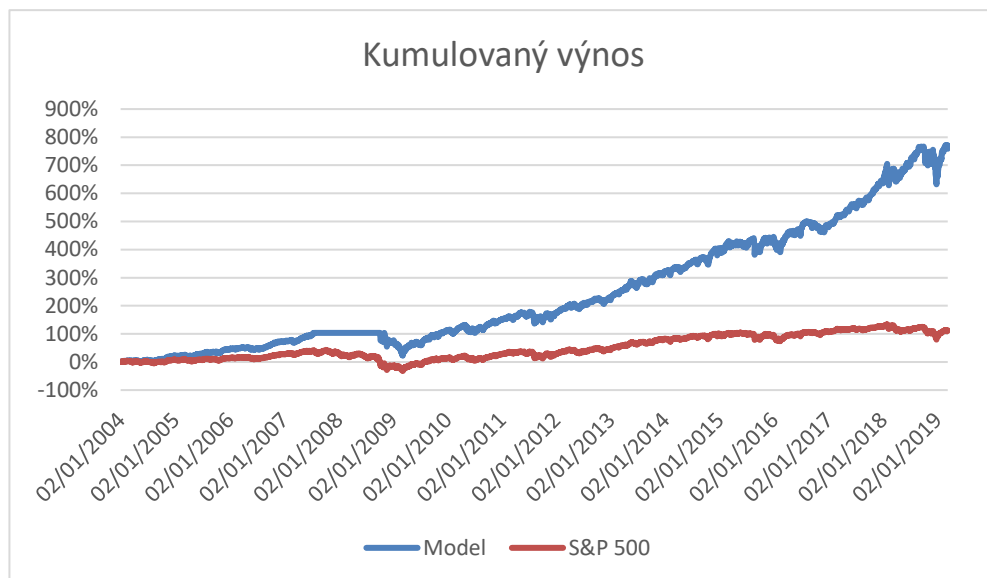
Zdroj: Vlastní zpracování v programu Qlik Sense

Posledním modelem je model, kterým nám určuje podíl aktiv v portfoliu. Nový prvkem je zde koláčový graf, který určuje za dané období. Tento graf dává největší smysl, pokud se vybere jeden konkrétní den, pak zobrazuje přesnou skladbu k danému dni, jinak ukazuje průměr za více dnů.

Nesmí chybět porovnání modelu s jeho benchmarkem, tedy indexem S&P 500, jak v grafické podobě, tak v podobě výnosu za dané období. Období je zde zvoleno na průběh roku 2001, můžeme si proto všimnout velmi světlé heatmapy a značně záporných zisků, z důvodu technologické krize.

### 4.3 Analýza výkonnosti modelu

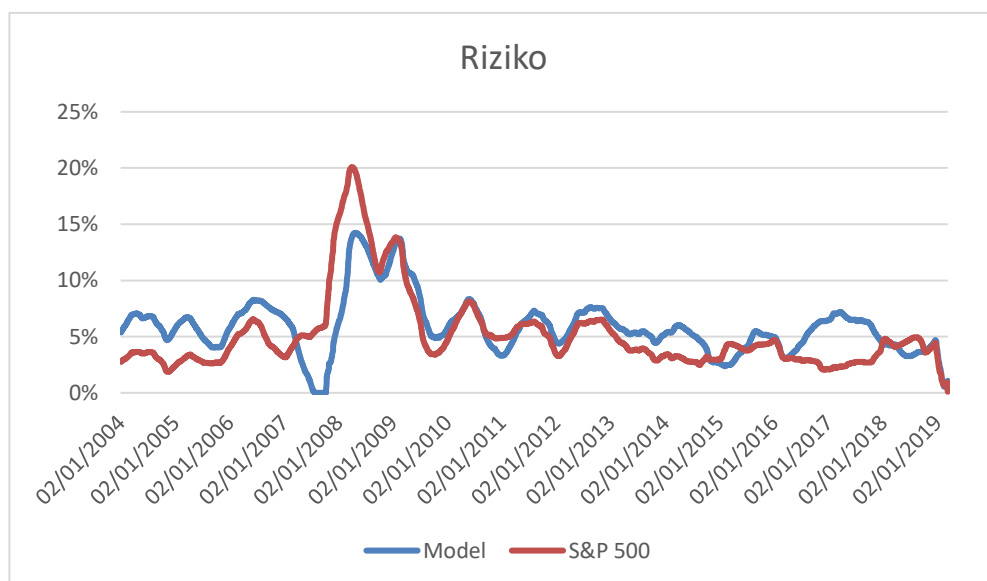
Poslední otázkou, co zbývá je, jestli model porazí svůj benchmark, tedy index S&P 500. Jak můžeme vidět na obrázku 38, model opravdu svůj benchmark porazil.



**Obrázek 38:** Porovnání kumulovaného výnosu od roku 2004

*Zdroj: vlastní zpracování*

Důležité také je zohlednit, s jakým rizikem tohoto výnosu dosáhl. Ač riziko (na obrázku 39) v průběhu času značně kolísá, nelze tvrdit, že by bylo signifikantně vyšší, než v případě indexu S&P 500.



**Obrázek 39:** Porovnání volatility od roku 2004

*Zdroj: vlastní zpracování*

## ZÁVĚR

Cílem této práce bylo vytvořit model s využitím externích dat. Tento model byl vytvořen za pomoci programovacího jazyku Python. Data do tohoto modelu byla automatizovaně posbírána a nahrána za pomoci platformy UI Path a nakonec byl model vizualizován v aplikaci Qlik Sense.

Výsledný model byl vytvořen skombinováním dvou modelů, kdy každý modeloval rizikové faktory na jiné úrovni, jeden na úrovni jednotlivých akciových titulů a druhý na úrovni sektorů. Pro účel konstrukce těchto modelů byla využita logistická regrese, rozhodovací stromy a simulace Monte Carlo.

V práci je mimo jiné čtenář seznámen s metodami datové analýzy, s programovacím jazykem Python a nastupujícím trendem automatizace. Vedle toho byly zmíněny metody technické analýzy.

Závěrem by bylo vhodné dodat, že ač technologie prostupuje do světa financí, čím dál více, lidský faktor zde zcela nahradit nelze. Vždy bude potřeba expertů, kteří učiní rozhodnutí, na základě vlastních zkušeností a na základě doporučení systému.

## POUŽITÁ LITERATURA

- [1] BERNSTEIN, P. L. *The curious history of stock prices and interest rates*. 1979.
- [2] *Data science & big data analytics: discovering, analyzing, visualizing and presenting data*. Indianapolis, Indiana: John Wiley & Sons, 2015. ISBN 978-1-118-87613-8.
- [3] FANTA, Jiří. *Psychologie, algoritmy a umělá inteligence na kapitálových trzích*. Praha: Grada, 2001. Finance. ISBN 80-247-0024-7
- [4] FEDERAL RESERVE BANK OF ST. LOUIS *Commercial and Industrial Loans, All Commercial Banks*[online]. 2019 [cit. 2019-04-27]. Dostupné z: <https://fred.stlouisfed.org/series/BUSLOANS>
- [5] FEDERAL RESERVE BANK OF ST. LOUIS *M2 Money Stock* [online]. 2019 [cit. 2019-04-27]. Dostupné z: <https://fred.stlouisfed.org/series/M2>
- [6] FOTR, Jiří a Jiří HNILICA. *Aplikovaná analýza rizika ve finančním managementu a investičním rozhodování*. 2., aktualiz. a rozš. vyd. Praha: Grada, 2014. Expert. ISBN 978-80-247-5104-7.
- [7] GOLDHUB *Gold prices* [online]. 2019 [cit. 2019-04-27]. Dostupné z: <https://www.gold.org/goldhub/data/gold-prices>
- [8] MÁRTON, Peter a Norbert ADAMKO. *Praktický úvod do modelovania a simulácie*. Žilina: EDIS - vydavateľ'stvo ŽU, c2011. Vysokoškolské učebnice. ISBN 978-80-554-0387-8.
- [9] MELOUN, Milan a Jiří MILITKÝ. *Statistical data analysis: a practical guide : complete with 1250 exercises and answer key on CD*. New Delhi: Woodhead Publishing India, 2011. Woodhead Publishing India in materials. ISBN 978-0-85709-109-3.
- [10] MEERSCHAERT, Mark M. *Mathematical modeling*. 4th ed. Waltham: Academic Press, 2013. ISBN 978-0-12-386912-8.
- [11] MUSÍLEK, Petr. *Trhy cenných papírů*. 2., aktualiz. a rozš. vyd. Praha: Ekopress, 2011. ISBN 978-80-86929-70-5.
- [12] OECD Data: Consumer confidence index (CCI) [online]. 2019 [cit. 2019-04-27]. Dostupné z: <https://data.oecd.org/leadind/consumer-confidence-index-cci.htm>

- [13] PETR, Pavel. *Metody Data Miningu*. Pardubice: Univerzita Pardubice, 2014. ISBN 978-80-7395-872-5.
- [14] *Python* [online]. 2019 [cit. 2019-04-07]. Dostupné z: <https://www.python.org/>
- [15] *Python Brochure Vol.1* [online]. 2014, [cit. 2019-04-07].
- [16] REJNUŠ, Oldřich. *Finanční trhy*. 3., rozš. vyd. Ostrava: Key Publishing, 2011. Ekonomie. ISBN 978-80-7418-128-3.
- [17] *Slick Charts: S&P 500 Companies by Weight* [online]. 2019 [cit. 2019-04-27]. Dostupné z: <https://www.slickcharts.com/sp500>
- [18] STATSOFT, INC. (2013). *Electronic Statistics Textbook*. Tulsa, OK: StatSoft. WEB: <http://www.statsoft.com/textbook/>
- [19] STEWART, John M. *Python for scientists*. Cambridge: Cambridge University Press, c2014. ISBN 978-1-107-68642-7.
- [20] SWEDROE, Larry E. a RC BALABAN. *Úspěšný investor: odhalte 77 investičních omylů*. Brno: BizBooks, 2012. ISBN 978-80-265-0052-0.
- [21] ŠPAČEK, Miroslav. *Pravděpodobnostní přístupy k analýze rizik investičních projektů a jejich využití v praxi*. Plzeň: Nava, 2014. ISBN 978-80-7211-472-6.
- [22] ŠTÝBR, David, Petr KLEPETKO a Pavlína ONDRÁČKOVÁ. *Začínáme investovat a obchodovat na kapitálových trzích*. Praha: Grada, 2011. Finance pro každého. ISBN 978-80-247-3648-8.
- [23] U.S. DEPARTMENT OF THE TREASURY *Daily Treasury Yield Curve Rates* [online]. 2019 [cit. 2019-04-27]. Dostupné z: <https://www.treasury.gov/resource-center/data-chart-center/interest-rates/pages/TextView.aspx?data=yieldAll>
- [24] VAN DER AALST, Wil M. P., Martin BICHLER a HEINZL. *Robotic Process Automation* [online]. 14.5.2018, 269-272 [cit. 2019-04-07]. Dostupné z: <https://link.springer.com/content/pdf/10.1007%2Fs12599-018-0542-4.pdf>
- [25] VESELÁ, Jitka. *Investování na kapitálových trzích*. 2., aktualiz. vyd. Praha: Wolters Kluwer Česká republika, 2011. ISBN 978-80-7357-647-9.

[26] WILLCOCKS, L., M. LACITY a A. CRAIG. *The IT function and robotic process automation* [online]. 10/2015, 3-9 [cit. 2019-04-07]. Dostupné z: [http://eprints.lse.ac.uk/64519/1/OUWRPS\\_15\\_05\\_published.pdf](http://eprints.lse.ac.uk/64519/1/OUWRPS_15_05_published.pdf)

[27] YAHOO FINANCE *Bussines Finance, Stock Market, Quotes, News* [online]. 2019 [cit. 2019-04-27]. Dostupné z: <https://finance.yahoo.com/>