Univerzita Pardubice
Fakulta ekonomicko-správní
Ústav systémového inženýrství a informatiky

Detekce finančních podvodů metodami strojového učení

Bc. Nika Beranová

Diplomová práce
2018

Univerzita Pardubice
Fakulta ekonomicko-správní
Akademický rok: **2017/2018**

ZADÁNÍ DIPLOMOVÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení:	**Bc. Nika Beranová**
Osobní číslo:	**E15680**
Studijní program:	**N6209 Systémové inženýrství a informatika**
Studijní obor:	**Informatika ve veřejné správě**
Název tématu:	**Detekce finančních podvodů metodami strojového učení**
Zadávající katedra:	**Ústav systémového inženýrství a informatiky**

Zásady pro vypracování:

Cílem práce pro detekci finančních podvodů metodami strojového učení je zabezpečení sběru dat, popis získaných dat, charakteristika vybrané metody strojového učení, navržení modelu pro detekci finančních podvodů, jeho verifikace a provedení porovnání výsledků zvolených metod.

Osnova:
- Sběr a předzpracování dat
- Charakteristika vybrané metody strojového učení
- Navržení modelu pro detekci finančních podvodů
- Verifikace, porovnání výsledků zvolených metod

Rozsah grafických prací:

Rozsah pracovní zprávy: **cca 60 stran**

Forma zpracování diplomové práce: **tištěná/elektronická**

Seznam odborné literatury:

[1] WITTEN, I.H., FRANK, E., HALL, M.A. Data Mining: Practical Machine Learning Tools and Techniques. Amsterdam: Morgan Kaufmann, 2011. 665 s. ISBN 978-0-12-374856-0.
[2] ALPAYDIN, E. Introduction to Machine Learning. London: MIT Press, 2009. 579 s. ISBN 978-0-262-01243-0.
[3] MITCHELL, T. Machine Learning. New York: McGraw-Hill, 1997. 432 s. ISBN 0070428077.
[4] WEISS, S. M., INDURKHYA, N., ZHANG, T. Fundamentals of Predictive Text Mining. New York: Springer, 2010. 226 s. ISBN 978-1849962254.
[5] MINER, G. Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications. Amsterdam: Elsevier, 2012. 1000 s. ISBN 978-0-12-3870117.

Vedoucí diplomové práce: **doc. Ing. Petr Hájek, Ph.D.**
Ústav systémového inženýrství a informatiky

Datum zadání diplomové práce: **1. září 2017**

Termín odevzdání diplomové práce: **30. dubna 2018**

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

L.S.

doc. Ing. Pavel Petr, Ph.D.
vedoucí ústavu

V Pardubicích dne 1. září 2017

PROHLÁŠENÍ

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

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

Beru na vědomí, že v souladu s § 47 b 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. 2018 Bc. Nika Beranová

PODĚKOVÁNÍ:

Na tomto místě bych chtěla poděkovat svému vedoucímu práce panu doc. Ing. Petru Hájkovi, Ph.D. za jeho odbornou pomoc, cenné rady, trpělivost a poskytnuté materiály při zpracování této diplomové práce. Dále bych chtěla poděkovat své rodině za pochopení a podporu při studiu.

ANOTACE

Diplomová práce se zabývá problematikou detekce finančních podvodů metodami strojového učení. Základní koncept pojednání je situován do čtyř hlavních oblastí, a sice kdy jsou v první části shrnuty základní problematiky finančních podvodů z hlediska pojmů, dělení, preventivních opatření a detekčních technik. Druhá část je zaměřena na shrnutí výsledků z reportu SEC z roku 2010-2015 a třetí část na teoretické základy strojového učení a vybrané algoritmy. Poslední část práce je věnována popisu sběru dat a jejich přípravě k měření, přičemž v poslední kapitole jsou získané výsledky porovnány a statisticky vyhodnoceny.

KLÍČOVÁ SLOVA

strojové učení, finanční podvody, detekce, finanční výkazy, firmy

TITLE

Detecting Financial Fraud by Machine Learning Methods

ANNOTATION

This Master's thesis deals with the issue of detection of financial frauds by machine learning. The basic concept of the treatise is situated in four main areas, although the first part summarizes the basic problems of financial frauds in terms of concepts, divisions, preventive measures and detection techniques. The other part is focused on summarizing the results from the SEC report from 2010-2015 and the final part on the theoretical foundations of machine learning and selected algorithms. The final part of the thesis is devoted to the description of data collection and its preparation for measurement, whereas in the last chapter are the obtained results compared and statistically evaluated.

KEYWORDS

machine learning, financial fraud, detection, financial statement, companies

OBSAH

SEZNAM TABULEK

SEZNAM OBRÁZKŮ

SEZNAM ZKRATEK

ACFE	Association of Certified Fraud Examiners
AUC	Area Under Curve
COSO	Committee of Sponsoring Organizations
ISA	International Standards on Auditing
MLP	Multilayer Perceptron
NASDAQ	National Association of Securities Dealers Automated Quotations
NYSE	New York Stock Exchange
RBF	Radial Basic Function
ROC	Receiver Operating Characteristic
SEC	Securities and Exchange Commission
SEED	Securities Enforcement Empirical Database
SMO	Sequential Minimal Optimization
SVM	Support Vector Machines
TDIDT	Top Down Induction of Decision Trees
USD	United States Dollar

ÚVOD

Ačkoliv se může zdát, že s nástupem moderních technologií se povědomí společnosti o přítomnosti finančních podvodech zvýšilo, problematika výskytu podvodů není pouze záležitostí dnešní doby. Obecně lze říci, že k datování finančních podvodů dochází v důsledku existence vzniku finančních prostředků. Z historických milníků nelze opomenout období 20. století, kde je jistě velmi známým představitelem finančních podvodů Charles Ponzi, jehož jméno vstoupilo do povědomí širší veřejnosti v podobě schématu nesoucí jeho jméno. V souvislosti s nedávnou minulostí pak lze zmínit i případ Bernarda Madoffa, souzeného za nelegální obohacení se právě díky výše jmenovanému Ponziho schématu. Tyto a mnohé další případy pak pouze poukazují na setrvávající existenci finančních podvodů v průběhu několika období.

Finančních podvody nejsou ovšem záležitostí pouze dílčích osob. Americká komise pro cenné papíry a burzy (SEC) v průběhu let 2010-2015 vedla proti veřejným obchodním společnostem v průměru 735 vynucovacích akcí ročně, přičemž nejčastějším typem tvrzení obviňující společnosti bylo pochybení emitenta v souvislosti s porušením zákona o cenných papírech (Choi, 2016, s. 3-5). Odlišný, avšak v posledních letech také hojně se vyskytující typ podvodných praktik, bývá i zkreslování údajů ve finančním výkaznictví, přičemž důsledkem tohoto konání může být způsob zvýšení cen akcií nebo získání bankovních výpůjček (Huang, 2016, s. 1-2).

Důvodů, proč se jednotlivci dopouští finančních podvodů, je ovšem spousta. Častou příčinou bývá motiv finančního zisku, jenž může mít podobu platů nebo bonusů, dosáhne-li společnost stanovených cílů. Dalším podmětem může být také tlak, který je na jedince vyvíjen z vyšších pozic managementu nebo plyne z osobních problémů jedince. Naopak jinou postačující příčinou může být i touha po přesvědčení věřitelů nebo investorů (Zack, 2009, s. 5-6). Ačkoliv se společnosti podvodných praktik účastní s vidinou zvýšení svého zisku, případné odhalení má na vývoj společnosti dopad nepříznivý. Studie (Beasley, 1999, s. 37-40) uvádí, že po zveřejnění informací o podvodu se u 204 společností, jež se dopustily klamání v oblasti finančního výkaznictví, projevily následující efekty: 36 % společností ukončilo svou činnost nebo zbankrotovalo, 15 % společností prodalo velkou část svých aktiv a u 21 % společností došlo ke zrušení akcií z národní burzy cenných papírů. Obdobná studie poté ukazuje, že po zveřejnění podvodu klesla u dvaceti společností průměrná cena akcií v průměru až o 58 %.

Způsob detekce finančních podvodů je ovšem záležitostí nesnadnou. Auditorská činnost se stává obtížnou v důsledku možného nedostatku potřebných znalostí auditorů, což může vyplývat z méně častých výskytů konkrétních typů podvodů nebo možného ovlivňování jedinců z řad auditorů ze strany manažerských pozic (Huang, 2016, s. 5). Dalším problémem může být i konflikt zájmů mezi vlastním zájmem auditora a jeho profesní odpovědností (Moore, 2006, s. 10). Tradiční způsob odhalování finančních podvodů se tak může jevit, v důsledku výše uvedených příčin, jako nedostatečný a mnohdy i nákladný. Stále častějším způsobem pak bývají k analýze problematiky detekce finančních podvodů využívány metody z oblastí umělé inteligence (West, 2016, s. 47-66; Ngai, 2011, s. 559-569).

Cílem této práce je pomocí metod strojového učení klasifikovat data společností, které se dopustily nebo nedopustily podvodného jednání v období let 2012–2015. V rámci splnění cíle bude třeba splnit dílčí cíle, tj. zabezpečení sběru dat, popis dat, charakteristika vybraných metod strojového učení, navržení modelu pro detekci finančních podvodů, verifikace modelu na reálných datech a provedení statistického porovnání výsledků zvolených metod.

Práce bude rozdělena do šesti hlavních kapitol. V první části práce bude nahlíženo na problematiku finančních podvodů z obecného pohledu. Obsahem této kapitoly bude vysvětlení základních definic a pojmů týkajících se výkladu finančních podvodů a jejich základních dělní. Druhá kapitola bude věnována možnostem prevence a detekce. Třetí část bude zaměřena na shrnutí výsledků z reportu Komise pro cenné papíry a burzy z období let 2010–2015. Čtvrtá kapitola bude již zaměřena na teoretický obsah tématu strojového učení. Rozebrána zde tak bude základní koncepce strojového učení společně s popisem vybraných algoritmů. Pátá kapitola bude věnována souhrnným informacím týkajících souboru dat a jejich předzpracování k další části práce. Náplň této kapitoly tak bude směřována k popisu sběru dat, jejich zhodnocení a návrhu klasifikačního modelu. Poslední část bude zaměřena již na samotné provedení experimentů a shrnutí získaných výsledků porovnáním klasifikačních metod.

1. FINANČNÍ PODVODY

Obsah této kapitoly bude zpočátku věnován celkové problematice podvodu, s pozdějším zaměřením na podvody finanční. Následující část práce tak bude věnována obecným definicím podvodného jednání a grafickému vyjádření podvodu v podobě trojúhelníku podvodu, diamantu podvodu a modelu M.I.C.E. V navazujících částí budou popsány základní možnosti a způsoby klasifikace zaměřené již na podvody finanční.

1.1. Vymezení pojmu

V důsledku širokého spektra jednotlivých definic a termínů vymezujících formulaci podvodu, zde budou uvedeny definice znázorňující shodné i odlišné aspekty koncepce podvodu a klamavého finančního jednání.

V souvislosti s obecným právem je podvod tvořen čtyřmi prvky, jejichž význam tvoří záměrně zkreslená prohlášení, vědomost o nepravdivé výpovědi, spoléhání na nepravdivou výpověď a odškodné vyplývající ze spolehnutí se na nepravdivé prohlášení oběti (Kranacher, 2011, s. 2-3). Encyclopaedia Britannica (Fraud, 2017) uvádí formulaci podvodu jako *„úmyslné zkreslení skutkového stavu s cílem zbavit někoho cenného držení“*. Dále pak uvádí, že *„ačkoliv je podvod trestným činem sám o sobě, častěji je prvkem zločinů, v souvislosti se ziskem peněz pod falešnou záminkou nebo předstíranou identitou“*.

Odlišnou definici uvádí mezinárodní auditorský standart ISA 240 (2004, s. 3), dle kterého je podvod definován jako „*úmyslný čin, jehož se dopustí jedna nebo více osob z řad vedení, zaměstnanců, osob pověřených řízením nebo třetích stran a který má formu klamu za účelem získání neoprávněné nebo protiprávní výhody*“.

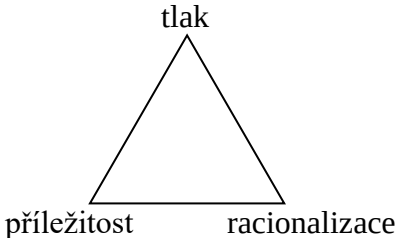
V nejširším pojetí lze tedy tvrdit, že podvodem může být jakýkoliv zločin, využívající klamavých technik jako své hlavní preference (Kranacher, 2011, s. 3), a tedy v návaznosti na výše uvedené formulace lze vymezení finančního podvodu vyjádřit jako *„úmyslné použití nezákonných metod nebo postupů za účelem získání finančního zisku“* (West, 2016, s. 47).

Jak je možné vidět z výše uvedených definic, ve všech případech je zmíněné jednání definované jako úmyslné. Nutno tedy zdůraznit, že v kontextu s podvody je rozlišováno mezi jednáním způsobujícím úmyslnou újmu a jednáním neúmyslným, vznikajícím v důsledku neúmyslné nesprávnosti označované jako chyba, která předmětem podvodu není (Young, 2014, s. 4-5).

1.2. Schématické vyjádření

1.2.1. Trojúhelník podvodu

Prvotní schématický zápis podvodného jednání formuloval v roce 1953 kriminalista Dr. Donald Cressey. Jedná se o tzv. trojúhelník podvodu, jehož grafické znázornění je možné vidět níže (Obrázek 1). Cressey se ve svém výzkumu, započatým v roce 1950, zabýval otázkou, proč lidé páchají podvodné úkony, a tedy i jaké jsou samotné příčiny k porušení něčí důvěry. Na základě dotazovacího výzkumu skupiny testovaných osob, odsouzených za zpronevěru, došel k závěru, že jedinec páchající podvod je ovlivněn společnou existencí tří faktorů: *příležitostí*, *racionalizací* a *tlakem* (Kassem, 2012, s. 191-192).



Obrázek 1: Trojúhelník podvodu

Zdroj: upraveno dle (Kassem, 2012, s. 192)

Příležitost ke spáchání podvodného činu, je dána znalostí konkrétního místa, kde jest činnost páchána a nízké pravděpodobností odhalení podvodu, přičemž tato pravděpodobnost je dána důsledkem slabé vnitřní kontroly (Dorminey, 2012, s. 557-558). Schutchter (2016, s. 111) v tomto ohledu dále jmenuje tři nejvýraznější nedostatky, které vnitřní kontrola nejčastěji postrádá, a které tak dávají k podvodnému jednání největší příležitost. V prvních dvou případech autor zmiňuje absenci sledovacích zařízení a bezpečnostních opatření. Další problém pak vidí v příčině nedostatečných schvalovacích systémů opravňující jedince k činnostem.

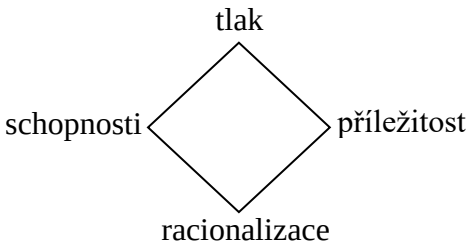
V případě faktoru racionalizace jde o vlastní přesvědčení jednice, že skutky jeho konání jsou správné a nepředstavují tak pro něj trestný čin, za který by měl být souzen (Dorminey, 2012, s. 558). Goldmann (2010, s. 15) dále doplňuje, že pachatel své jednání může zpočátku vnímat jako špatné, ale proces ospravedlnění nabyde podstaty až po úmyslu navrácení získaných aktiv pachatelem.

Motiv tlaku je vytvářen jedincovou neschopností sdílet své problémy, přičemž tato neschopnost může být dána jedincovým egem nebo odmítnutí pomoci od okolí (Dorminey, 2012, s. 558). Vyvolaný tlak vedoucí k trestnému jednání tak může být dle Goldmanna (2010,

s. 14) důsledkem dlouhodobé nezaměstnanosti, neschopností splácet dluhy nebo jinými finančními problémy, které donutí jedince k motivu páchání trestné činnosti.

1.2.2. Diamant podvodu

Rozšiřujícím modelem trojúhelníku podvodu je model nazývaný diamant podvodu (Obrázek 2) představený v roce 2004 dvojicí Wolfe a Hermanson. Tento typ modelu vychází z původního Creesyho modelu a jak je možné vidět na níže uvedeném grafickém znázornění, je tento typ vzoru rozšířen o faktor *schopnosti* (Kassem, 2012, s. 194).



Obrázek 2: Diamant podvodu

Zdroj: upraveno dle (Kassem, 2012, s. 194)

Schopnost jedince páchající podvod je dle výše jmenované dvojice (Wolf a Hermanson) nedílným aspektem ke spáchání podvodu, neboť osoba bez patřičné způsobilosti a osobních rysů by k vedení trestného činu nemohla dospět (Dorminey, 2012, s. 564). Mezi charakteristické znaky jednotlivce páchající podvod zařadila dvojice tvůrců modelu viditelné vlastnosti, dle kterých je osoba spíše v pozici vyšší funkce organizace, má znalosti v účetních systémech, zná interních nedostatky v organizace, má jistotu v nepostihnutelnosti a je psychicky imunní vůči konání nezákonných činností (Kassem, 2012, s. 194).

1.2.3. M.I.C.E

V návaznosti na předchozí dva výše jmenované modely lze zmínit i třetí model uváděný pod zkratkou M.I.C.E (Money, Ideology, Ceorcion, Ego). Tento typ modelu byl navrhnut profesorkou Mary-Jo Kranacher v roce 2010 a poukazuje na detailnější rozšíření faktoru tlaku. Odlišný přístup pohledu na motivaci ke spáchání podvodu poukazuje na čtyři činitele, jimiž jsou *peníze*, *ideologie*, *donucení* a *ego* (Kassem, 2012, s. 194).

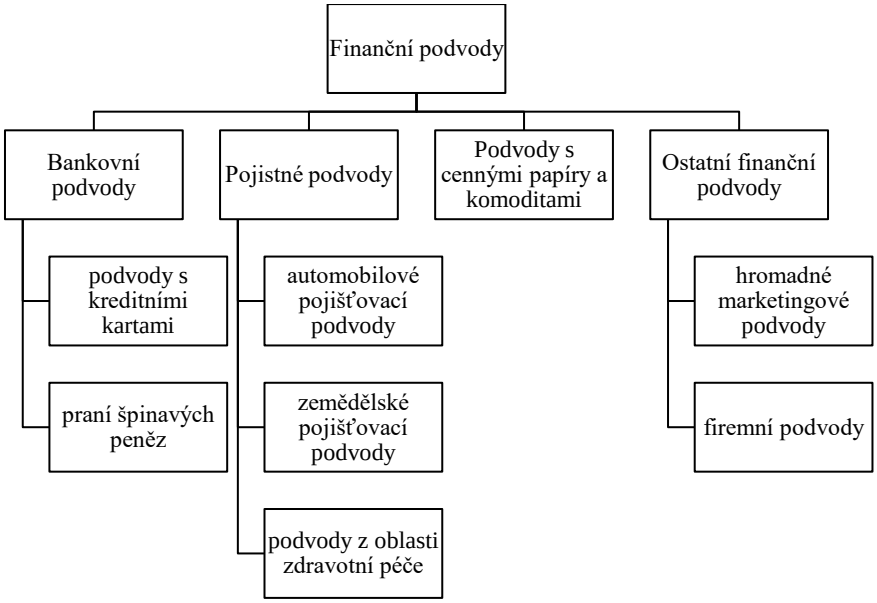
Dle Kranacher (2011, s. 205) je nejčastější motivací ke spáchání podvodu touha po penězích a ego jedince, které dle Kassem (2012, s. 194) může být motivem ke spáchání podvodu z důvodu obav ztráty společenského nebo rodinného postavení. Kranacher dále uvádí, že naopak nejméně motivujícím faktorem je vliv donucení, dle kterého může být osoba vystavena

podvodným aktivitám pod nátlakem jiné osoby. Motiv ideologie pak bývá častým příkladem podvodu v podobě daňových úniků či financování terorismu, avšak i zde je tento typ motivace ne příliš častým podnětem ke klamání (Dorminey, 2012, s. 563).

1.3. Klasifikace finančních podvodů

Možných způsobů dělení finanční podvodů je mnoho. Spousta autorů se neshoduje na konkrétním, jednotném členění a názory bývají v tomto směru mnohdy rozdílné. S ohledem na tématiku práce bude obsah této části nejdříve věnován obecnému dělení finančních podvodů a poté klasifikaci se zaměřením na podvody konané v podnikovém sektoru.

Na následujícím znázornění (Obrázek 3) je zachyceno členění do čtyř základních kategorií, jak jej uvádí Ngai (2011, s. 563). Autor dělí finanční podvody na podvody *bankovní*, *pojistné*, *podvody s cennými papíry a komoditami* a *podvody nezařaditelné*.



Obrázek 3: Dělení finančních podvodů

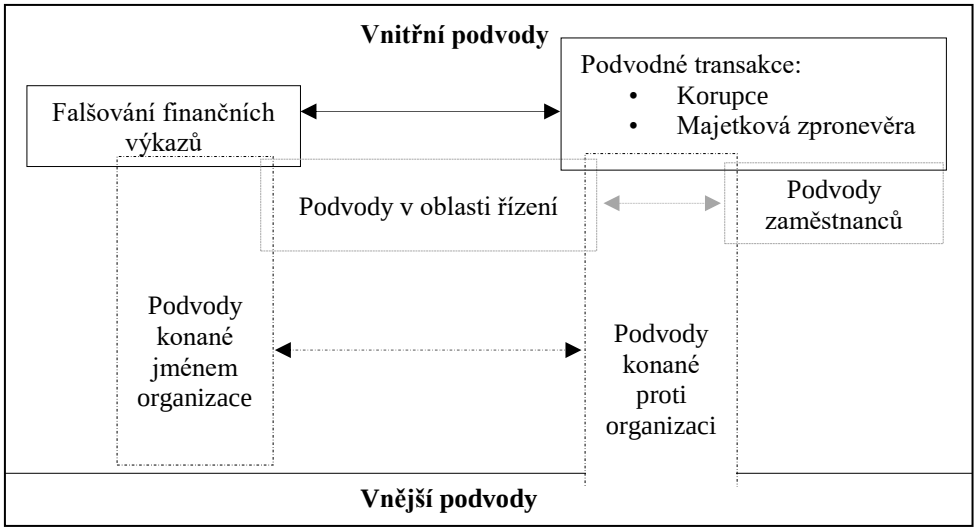
Zdroj: upraveno dle (Ngai, 2011, s. 563)

Bankovní podvody jsou typem trestných činů, během kterých dochází ke snaze oklamat nebo jinak podvést finanční instituce (Gottschalk, 2010, s. 448). K oklamání těchto institucí může dojít prostřednictvím podvodných prohlášení, jejichž snahou je odcizit finanční prostředky instituce, mající podobu cenných papírů, úvěrů nebo aktiv (Ngai, 2011, s. 561). Příkladem tohoto typu podvodu mohou být podvody s kreditními kartami, praní špinavých peněz nebo podvody hypotéční, na základě kterých dochází k úmyslné manipulaci s majetkovými dokumenty s cílem ovlivnit potenciální věřitele (West, 2016, s. 50-51).

Pojistné podvody mohou mít mnoho podob. Dle Ngai (2011, s. 561) se jedná o podvody, které mohou být páchány jakýmkoliv jedincem v průběhu celého procesu pojištění, přičemž příkladem těchto jedinců mohou být spotřebitelé, zaměstnanci pojišťovny nebo poskytovatele zdravotní péče. Typ tohoto podvodu zmiňuje také Gottschalk (2010, s. 445), který nazývá typ tohoto klamání podvodem spotřebitelským, ve kterém spotřebitelé záměrně zkreslují fakta s cílem získat pojistné plnění. Typickým zástupcem těchto činů, jak uvádí Ngai (2011, s. 561), jsou podvody páchané v automobilovém průmyslu nebo v oblasti zdravotní péče.

Podvody s cennými papíry a komoditami jsou podvody, ke kterým dochází v situaci, kdy je osoba na základě zkreslených informací oklamána, aby investovala do společnosti. Obsahem těchto podvodů mohou být podvodná schémata, podvody s hedgeovými fondy nebo zpronevěra (West, 2016, s. 50-51).

Sekce ostatních finančních podvodů pak může zahrnovat podvody z oblasti podniku nebo podvody z oblasti masového marketingu. Finanční podvody z oblasti masového marketingu jsou vedeny jako souhrnný název pro podvody konané přes komunikační média, jimiž mohou být činy páchané přes internet nebo telemarketing (Ngai, 2011, s. 562). Firemní podvody jsou podvody páchané z pohledu společností, přičemž na tyto podvody lze nahlížet buď ze strany typu podvodu nebo ze strany typu pachatele (O'Gara, 2004, s. 2). Na následujícím znázornění (Obrázek 4) je uvedeno vyjádření klasifikace dle autorky Jans Mieke (2010, s. 208), která tímto znázorněním zachycuje náhled na problematiku klasifikace korporátních podvodů, jejichž detailnější popis bude uveden v následujících kapitolách.



Obrázek 4: Klasifikace firemních podvodů

Zdroj: upraveno dle (Jans, 2010, s. 208)

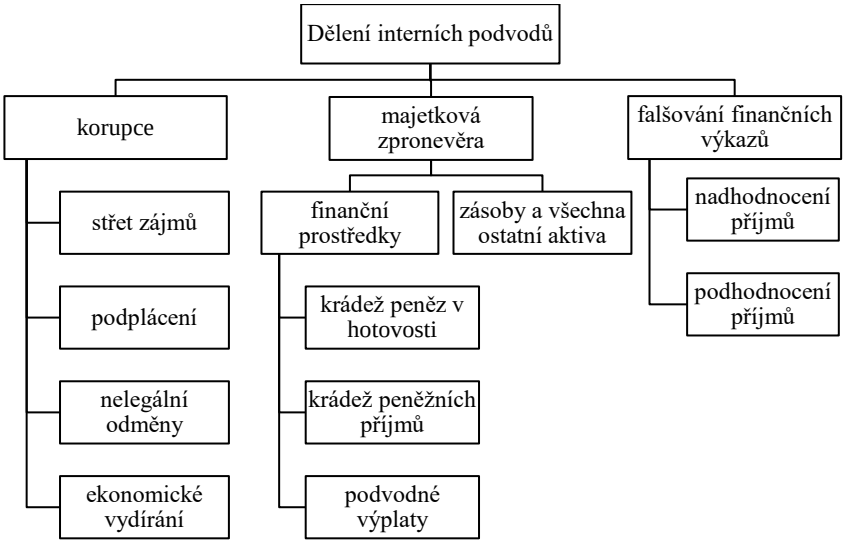
Z výše uvedeného schématického vyjádření si je možné povšimnout i jistých překrývání v rámci jednotlivých typů klasifikací. Tato překrývání souvisí s různými pohledy na rozdělení

podvodných jednání a jejich vzájemnými vztahy včetně pohledu na odlišení podvodu ze strany typu pachatele (Jans, 2013, s. 1666).

1.3.1. Vnitřní a vnější řešení

Klasifikace podvodů na podvody vnitřní a podvody vnější je jeden z nejčastějších a zároveň nejdůležitějších způsobů členění. Z pohledu vnějších nebo také externích podvodů, jsou pachateli trestných činů osoby vyskytující se mimo organizaci, přičemž se jedná se o tzv. třetí strany, jejichž jednání je vedeno proti dané společnosti. Těmito třetími stranami mohou být prodejci, dodavatelé nebo kontraktoři. Vnitřní podvody jsou naopak vykonávány osobami vyskytujícími uvnitř dané struktury organizace a mohou jimi být řadoví zaměstnanci nebo manažeři dané společnosti (Cruz, 2015, s. 23; Jans, 2010, s. 206).

Asociace certifikovaných vyšetřovatelů podvodů (ACFE) v souladu s vnitřními podvody zavádí systém klasifikace těchto jednání, pod označením podvodů pracovních nebo zaměstnaneckých, které uvádí do schématu (Obrázek 5) nazývaném také jako strom podvodů. V rámci této klasifikace, definuje ACFE tři základní typy interních podvodů[1], jimiž jsou *korupce*, *majetková zpronevěra* a *falšování finančních výkazů* (ACFE, 2016, s. 10-11).



Obrázek 5: Klasifikace interních podvodů

Zdroj: upraveno dle (ACFE, 2016, s. 11)

Význam korupce definuje Kranacher (2011, s. 4) jako situaci, kdy daná osoba zneužije svého postavení v transakčních obchodech s cílem dosáhnutí vlastního obohacení se na úkor osoby jiné. Mezi běžné příklady tohoto jednání je možné zařadit střet zájmů nebo přijímání úplatků.

[1] Podrobnější schématický zápis je uveden v příloze A

Nbaham (2009, s. 18) v souvislosti s podplácením zmiňuje dva směry tohoto jednání, a sice podplácení v podobě *bid riggingu*, kdy zaměstnanec společnosti zvýhodní pozici dodavatele ve výběrovém řízení a podplácení v podobě nezákonné provize, v angličtině označované jako *kickback*, kde tato provize může vzniknout v situaci, kdy dodavatel předloží zaměstnanci zkreslené fraktury, ze strany zaměstnance dojde k záměrné ignoraci těchto faktur, čímž je dodavatelem zaměstnanci zpětně vyplacen finanční podíl.

Za zpronevěru majetku je označován podvod, ve kterém se zaměstnanci dopouští krádeže majetku organizace nebo jej zneužijí ve svůj prospěch. Mezi běžné případy těchto podvodů patří mzdové podvody nebo odcizení zásob (Kranacher, 2011, s. 4).

Třetím základním typem interních podvodů je podvod týkající se falšování finančních výkazů. Finanční výkazy jsou dokumenty odrážející finanční stav společnosti, a tedy z obsahu těchto dokumentů je patrno, zda je daná organizace v kladné nebo záporné ekonomické situaci. Tento finanční stav je daný účetní závěrkou, kterou jsou společnosti povinny zveřejňovat každý rok a každé čtvrtletí. Obsah těchto uzávěrek, které mohou obsahovat rozvahy, výkazy zisku a ztráty nebo výkazy peněžních toků, jsou následně důležité pro akcionáře, kteří rozhodují o svých investicích do společnosti a pro banky, rozhodující o poskytnutí půjčky (Ravisankar, 2011, s. 491). Mezi typické příklady falšování finančních výkazů patří nadhodnocení příjmů nebo podhodnocení závazků a výdajů (Kranacher, 2011, s. 4).

Zde je možné si povšimnout rozdílnosti mezi klasifikací interních podvodů uváděné dle ACFE a uvedeným znázorněním dle Janse. Příčina této odlišnosti spočívá v dalším možném způsobu členění interních podvodů odlišující podvody transakčních charakterů, kde je smyslem odcizení organizačních aktiv, od podvodů konaných v rámci vykazování falešných zpráv, jejichž cílem je zlepšit finanční obraz společnosti (Jans, 2013, s. 1666).

1.3.2. Řešení s ohledem na organizaci

Způsob této klasifikace vyvstává z otázky, zda jsou podvody konány proti dané společnosti nebo zda pachatel jedná v souladu s danou společností a činí tak podvod jejím jménem.

Podvody konané jménem organizace jsou nejčastěji konány osobami z vedoucích nebo manažerských pozic instituce, jejichž konání vede k vytvoření zkresleného ekonomického stavu dané společnosti. Tento stav je poté důležitý pro akcionáře, věřitele nebo regulační orgány, na jejichž rozhodnutí může záviset budoucí stav organizace. Mezi příklady těchto podvodů může patřit nadhodnocování příjmů nebo podhodnocování výdajů a závazků (Singleton, 2006, s. 20). V rámci toho dělení je třeba zmínit i osobní prospěch plynoucí

z vykonaného podvodu, který zmiňuje (Dvořáček, 2003, s. 131). Oproti podvodům, které jsou vedeny proti organizaci, zde mají pachatelé užitek nepřímý, neboť forma osobního prospěchu vzniká až se zlepšením ekonomické situace organizace.

V případě podvodů vedených proti organizaci pak může být užitek pachatelů nepřímý i přímý, neboť pachatel může jednat sám za sebe nebo tak může činit prostřednictví jiné organizační struktury, mající podobu například v konkurenční společnosti. Singleton (2006, s. 20) dále uvádí, že tyto podvody slouží na rozdíl od podvodů konaných jménem společnosti pouze pro pachatele, a ne pro organizaci. Představiteli těchto podvodů tak mohou být prodejci, dodavatelé nebo kontraktoři. Mezi typické příklady těchto podvodů může patřit zpronevěra nebo krádež aktiv podniku.

S ohledem na výše uvedené znázornění si je možné povšimnout, že podvody konané ku prospěchu společnosti zasahují celou svou částí do oblasti vnitřních podvodů, naproti tomu podvody páchané ke škodě organizace zasahují i do oblasti podvodů vnějších. Tento poznatek souvisí s formulací vnějších korporátních podvodů, jejichž konání je v podstatě forma podvodů vedených proti organizaci (Jans, 2013, s. 1666).

1.3.3. Řešení s ohledem na pracovní pozici

Rozlišení těchto typů finančních podvodů je odvozeno z podstaty, zda se pachatelem podvodu stává pracovník vyskytující se spíše ve vyšších pozicích dané organizace, nebo zda je pachatelem řadový zaměstnanec. Z této klasifikace je tedy vidno, že osobou konající podvod může být jakýkoliv člen organizace, ovšem důvod motivu nebo použité nástroje mohou být v konečném výsledku rozdílné (Singleton, 2006, s. 22).

V kontextu s podvody konanými v oblasti řízení a podvody konanými na nižších úrovních zaměstnání je pozorovatelná souvislost mezi podvody páchanými v oblasti finančního výkaznictví a podvody s vlastnostmi transakčního charakteru. V tomto ohledu se předpokládá, že manažeři nebo jiní vedoucí pracovníci mohou páchat podvody v obou uvedených směrech, tedy jak ve vykazování falešných zpráv, tak v oblasti korupce nebo majetkové zpronevěry. Naproti tomu řadoví zaměstnanci budou spíše omezeni pouze na podvody transakční.

V rámci podvodů konajících se ve prospěch společnosti pak budou hlavními pachateli vedoucí pracovníci, avšak jejich rámec konání může zasahovat i do podvodů konaných proti společnosti. Naopak zaměstnanci na nižších úrovních budou spíše zasahovat pouze do podvodů mířených ke škodě organizace (Jans, 2013, s. 1666-1667).

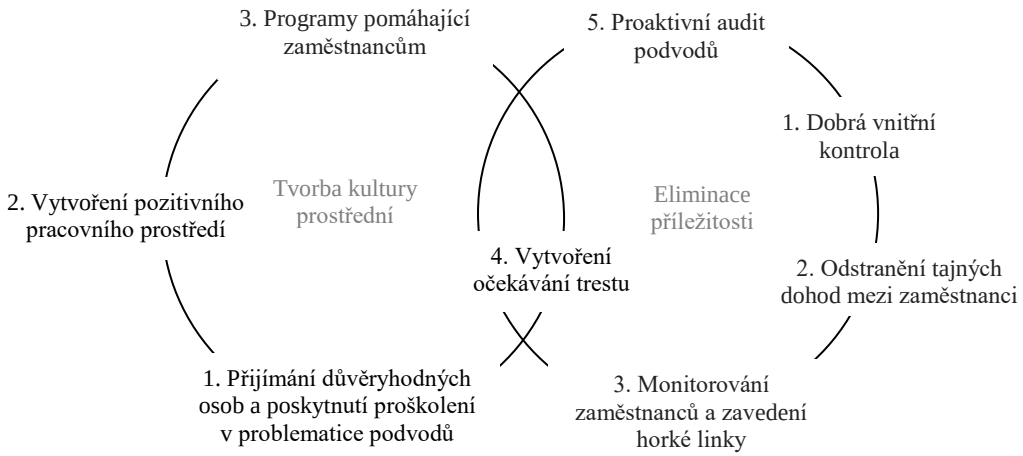
2. PREVENCE A DETEKCE

V první části kapitoly bude pojednáno o možných způsobech preventivních opatření napomáhajícím organizacím podvodným jednáním předcházet. Ohled zde bude kladen na tvorbu kultury prostředí a eliminaci faktoru příležitosti. V další části bude zmíněna významnost vnitřní kontroly a vlastnictví etického kodexu. Druhá část kapitoly pude věnována možným způsobům detekce, včetně přístupů pokročilých metod mající základ v umělé inteligenci.

2.1. Prevence proti podvodům

Zavedením strategických opatření proti výskytu podvodů je důležitým úkonem snižující riziko přítomnosti klamání a omezujícím prvkem výskytu, pokud k případným podvodům dojde. Důležitým aspektem pro předcházení podvodů jsou tak kontroly vycházející z vnitřních struktur organizace a kontroly vnějších obchodních vztahů mířené na zákazníky, dodavatele nebo jiné partnery tvořící s organizací obchodní poměr (Dubis, 2009, s. 19-20). Postupy, kterými je dosáhnuto prevence, by poté měly mít tři měřitelné cíle, a sice snižovat ztrátu způsobenou podvodem, odradit další podvody prostřednictvím proaktivních politik a zvyšovat pravděpodobnost včasného odhalení (Seetharaman, 2004, s. 1067).

Na následujícím zobrazení (Obrázek 6) jsou vyjádřena preventivní opatření dle Albrechta (2011, s. 120). Autor se zaměřuje na dvě hlavní strategie preventivních technik, a sice na předcházení vzniku podvodů prostřednictvím *tvorby kultury prostředí* a předcházení díky *eliminaci faktoru příležitosti*. Tyto dva směry a jejich části budou blíže rozebrány v následujících podkapitolách.



Obrázek 6: Prevence proti podvodům

Zdroj: upraveno dle (Albrecht, 2011, s. 120)

2.1.1. Tvorba kultury prostředí

Jak je možné si z výše uvedeného obrázku povšimnout, strategie tvorby kultury prostředí je vytvořena třemi okolnostmi, jimiž je podmínka náboru vhodných zaměstnanců a jejich následné proškolení, tvorba pozitivního pracovního prostředí a přítomnost programů pomáhajícím zaměstnancům zvládat jejich osobní zátěže.

Přijímání důvěryhodných osob do nového zaměstnaní, je důležitou součástí prevence, neboť ověření důvěryhodnosti může poukázat na předchozí případy podvodných jednání u jiných společností a je tedy pravděpodobné, že by daná osoba bez řádného prověření mohla své jednání v novém zaměstnání zopakovat (Martin, 1998, s. 13-14). K ověření uchazeče mohou zaměstnavatelé využívat formu životopisu nebo certifikace, přičemž tato forma prověření by měla být zaměstnavatelem i následně ověřena. Důležitou součástí po přijetí osoby je i její následné proškolení (Albrecht, 2011, s. 103-106). Prostřednictvím školení by tak nově přijatí zaměstnanci měli nabýt představu o formě přijatelného chování v organizaci a měli by být schopni identifikovat a vyhodnotit náznaky neetického chování zaměstnanců jiných. Informace poskytnuté členům organizace prostřednictvím přednášek, etického kodexu (kapitola 2.1.4) nebo jiných sdělovacích prostředků by tak měly zahrnovat informace o činech, které jsou organizací zakázané se spojením i s příslušnými zákony a předpisy, které jsou pro společnost platné. Dalším obsahem sdělovacích prostředků by měly být i informace o dopadu podvodu na společnost a jednotlivce, včetně poradenství, jak předcházet situacím, které by k případnému podvodu mohly vést (Lendez, 1999, s. 53).

Příznivé pracovní prostředí nemusí být vždy automatickou součástí podniku, avšak i tento faktor přispívá k prevenci. Albrecht (2011, s. 106-111) v tomto směru jmenuje tři okolnosti napomáhající organizaci vytvářet příznivou pracovní atmosféru za současného zmírnění rizika k podvodu. Mezi tyto tři podmínky autor řadí přítomnost *etického kodexu*, *politiku otevřených dveří* a *pozitivní personální a provozní politiky*, které zabraňují podvodům, mají-li zaměstnanci například mezi sebou pocit rovnosti, dostatečnou mzdu nebo dostatečné pracovní uznání. Otevřená komunikace poté umožňuje zabránit podvodům hned dvěma způsoby, a sice potlačuje klamání z důvodu minimalizace osobních tlaků a informuje vedoucí pracovníky o vlastnostech a příčinách těchto tlaků, díky čemuž mohou vedoucí pracovníci provést další kroky k preventivním opatřením. Jeager (2011, s. 69) dále doplňuje, že v souvislosti se získanými informacemi je vhodné, aby vedoucí pracovníci získané poznatky i mechanicky centralizovali, a sice z důvodu možného výskytu trendu, díky kterému lze řešit výskyt stejného typu problému i na více místech organizace.

Třetím faktorem vytvářející kulturu prostředí je prvek formálních programů napomáhající zaměstnancům zvládat osobní zátěže, mající původ například v hazardních hrách nebo návykových látkách. Tyto zátěže jsou často z pohledu jedince vnímány jako neřešitelná situace a mohou tak být postačujícím motivem pro budoucí trestnou činnost (Albrecht, 2011, s. 112). Programy pro pomoc zaměstnancům jsou ovšem využitelné pouze v případě, ovlivňuje-li nepříznivý stav jedince jeho pracovní výkon a lze tedy říci, že zavedením těchto programů organizace alternuje tradiční přístup varování nebo ukončení pracovního poměru (Johnson, 1985, s. 383).

2.1.2. Eliminace příležitosti

Další možností, jak zmírnit pravděpodobnost k výskytu podvodu, je neposkytnout pachatelům příležitost k jejich vykonání. Jak je možné vidět z výše uvedeného schématického vyjádření, je v tomto směru definováno pět možných způsobů k potlačení příležitosti, a sice pomocí opatření vnitřní kontroly, prostřednictvím potlačení tajných dohod mezi zaměstnanci, využitím monitoringu, vytvoření dojmu, že žádný podvod nebude tolerován a zavedením proaktivního auditorského přístupu.

Smyslem vnitřních kontrol je udržovat dohled nad správou informací a vývojem podniku, přičemž výsledkem jsou stavy, které by měly být v souladu se zavedenými organizačními normami (Dimitrijevic, 2015a, s. 35-36). Využitím tohoto přístupu je tak dán vedení společnosti nejvíce možný efektivní způsob, jak podvodům v organizaci předcházet, neboť díky tomuto postoji jsou identifikována slabá místa společnosti, která by měla být vedením napravena účinnější implementací kontrolních opatření. Pachatelé tak na základě těchto kontrol mohou ztrácet k podvodu příležitost, neboť jsou vedením oblasti podniku neustále sledovány (Seetharaman, 2004, s. 1067). Albrecht (2011, s. 113) v této souvislosti zmiňuje *Internal Control-Integrated Framework* vydaný Výborem pro sponzorské organizace Treadwayovy komise (COSO). S ohledem na obsáhlost tématiky bude této části prevence věnována podkapitola 2.1.3.

Zamezením tajných dohod mezi zaměstnanci a případně i mezi zaměstnanci a jinými subjekty spolupracujícími s danou organizací představuje důležitou součást preventivních opatření, neboť zaměstnanci dopouštějící se koluzních podvodů zpravidla páchají škody větší než jednotlivci. Důvodem větších ztrát je plošnější možnost utajení, a to zejména u složitých organizačních struktur, čímž je dána i časově náročnější detekce. Nejčastějšími typy podvodů tvořených tajnou dohodou jsou podvody z oblastí prodeje a nákupu (Albrecht, 2011, s. 116; Coenen, 2008, s. 14).

K předejití podvodům lze využít i formou monitoringu. Zaměstnanci, kteří spáchali podvod, pravděpodobně využijí zisk ve svůj prospěch nebo značně změní své chování, a tedy vnímáním těchto určitých jevů lze zpozorovat viditelné znaky, které mohou být pro manažery „červenými vlajkami“ (Albrecht, 2011, s. 117; Martin, 1998, s. 14). Martin se dále zaměřuje i na formy monitoringu na pozadí organizace. Kontroly by měli být prováděny periodicky a plošně nevyjímaje ani dlouhodobé zaměstnance, kteří jsou spíše náchylní k podvodu z důvodu znalosti fungování společnosti. Autor dále poukazuje na důležitost kontrol zejména v oblastech peněžních toků a v odděleních s výpočetní technikou (Martin, 1998, s. 14). S ohledem na informace získané prostřednictvím sledování je v důsledku prevence vhodné zavést i tzv. horkou linku (*whistle-blowing*). Tato linka může sloužit současným i bývalým zaměstnancům k podání oznámení, má-li oznamovatel podezření na přítomnost nečestného jednání. Tato oznámení jsou následně podávány osobám nebo organizacím, které mohou tato hlášení prošetřit (Near, 1985, s. 4). Důležitými prvky při zavedení horké linky je především zachování *anonymity* oznamovatele, *nestrannost* osob, jímž je stížnost podávána, *dostupnost*, při které zaměstnanci mají možnost využít několika různých prostředků k podání oznámení a *sledovanost* ze strany organizace nebo osob, jímž je stížnost podávána (Albrecht, 2011, s. 118).

Čtvrtý faktor eliminující příležitost k podvodu je forma prevence, vytvářející v jedincích pocit, že žádný z podvodů nebude tolerován. Nulová tolerance podvodu na jakékoliv úrovni organizace by se neměla obejít bez případného postihu, nevyjímaje ani podvody drobné, kde formou postihu může být pouze napomenutí. Závažnější podvody by se poté již neměli obejít bez trestního stíhání a ukončení pracovního poměru (Jarvis, 2002, s. 13).

Způsobem, jakým lze odstranit příležitost k podvodu, lze i prostřednictvím proaktivních auditů. Většina organizací se zaměřuje na audit reaktivní, kdy k identifikaci problému a jeho analýze dochází až poté, co se daný problém vyskytne. Naproti tomu audit proaktivní se zaměřuje na předvídání událostí, a tím i snížení rizika výskytu podvodu (Menkus, 1989, s. 30). Organizace využívající tento typ přístupu tak mohou v pachatelích příležitost k podvodu eliminovat, neboť zaměstnanci nad sebou získávají nepřetržitý pocit kontroly (Albrecht, 2011, s. 119).

2.1.3. Vnitřní kontrola

Dobrá vnitřní kontrola jako součást ochrany organizace proti výskytu podvodů je z pohledu prevence nejvíce uznávaný způsob, jak podvodům v organizacích předcházet (Albrecht, 2011, s. 113).

Vedení organizace, které je za fungování systému vnitřní kontroly odpovědné, využívá vnitřní kontrolu k získání přehledu nad určitými segmenty činností, aby mohlo lépe zhodnotit soulad mezi aktuálním stavem a rozvojovými cíli. Podstatou kontroly je tedy srovnání a výsledkem kontroly by mělo být tvrzení, že současný stav buď je anebo není v souladu se zavedenými normami. Na základě tohoto výroku lze tedy říci, že vnitřní kontrola neposkytuje jistotu ve vyloučení nesrovnalostí, ale zdůrazní pravděpodobnost selhání a očekává se tedy, že identifikuje lepší možnosti prevence (Dimitrijevic, 2015a, s. 36-38).

Nejvíce komplexní pohled na interní kontrolu poté zavádí Výbor pro sponzorské organizace Treadwayovy komise (COSO), který v roce 1992 zveřejnil základní standart pro hodnocení a zlepšení systému interní kontroly, vedený pod názvem *Internal Control-Integrated Framework*. Tento rámec umožňuje definovat společné vnímání vnitřní kontroly a zároveň vypomáhá vedením společností zlepšit nad podnikem dohled (Sinnett, 2003, s. 62; Dubis, 2009, s. 20). Definice kontroly, je tak pod tímto rámcem vedena jako *„proces, uskutečněný představenstvem, vedoucím a jiným personálem účetní jednotky, určený k poskytnutí přiměřené jistoty ohledně dosažení cílů v kategoriích: efektivních a účinných operací, spolehlivosti finančního výkaznictví a dodržování platných zákonů a předpisů“* (COSO, 2013, s. 3). Pro efektivní a účinný systém bylo následně komisí zavedeno pět obecně uznávaných prvků, jimiž jsou *kontrolní prostředí*, *posouzení rizik*, *kontrolní činnosti*, *informace a komunikace* a *monitoring* (Trenerry, 1999, s. 10). Těchto pět prvků je vzájemně propojeno a společně vytvářejí základ pro návrh prostředí, zabraňující pachatelům podvodnou činnost vykonávat (Dubis, 2009, s. 20).

Prvek kontrolního prostředí je základem pro všechny ostatní složky kontroly, neboť pro tyto složky zajišťuje základní strukturu. Jedná se o základ, ve kterém je vytvářena pracovní atmosféra organizace, přičemž obsahem této části by měla být integrita, etické hodnoty organizace nebo způsob jakým vedení řídí a rozvíjí své zaměstnance (Turner, 2009, s. 93). Albrecht (2011, s. 39-41) zde s ohledem na potlačení příležitosti k podvodu zmiňuje důležitost přístupu managementu ke svým zaměstnancům. Management by měl být pro zaměstnance příkladem čestného jednání a jeho komunikace by se neměla měnit v závislosti na okolnostech a situacích, neboť takovýto přístup vede k povzbuzení racionalizace. Dalším důležitým opatřením je i vhodný nábor zaměstnanců nebo přesné vymezení organizační struktury, díky které lze riziko výskytu podvodu také snížit, neboť je vedení přesně známa odpovědnost za každou obchodní činnost a případný podvod je tak možné snáze odhalit.

Druhý prvek představuje proces, posuzující vnější a vnitřní rizika organizace, neboť každá společnost nese určitou míru rizika, že se stane cílem podvodného jednání. Mezi faktory, které tato rizika způsobují, je možné zařadit změnu trhu nebo obrat zaměstnanců. Tyto okolnosti mohou nejen negativně pozměnit pravidelné činnosti společnosti, ale mohou ohrozit procesy organizace, a to včetně těch, které by měly společnost před podvody chránit (Turner, 2009, s. 94). Úkolem managementu by tedy mělo být tato rizika zhodnotit, určit míru dopadu těchto rizik na podnikání společnosti a sestavit odpovídající úroveň systému vnitřní kontroly (Trenerry, 1999, s. 13).

Třetí proces již reprezentují kontrolní činnosti, zajišťující existenci takových kontrol, která rizika pro dosažení organizačních cílů skutečně minimalizují (Trenerry, 1999, s. 14). Při určování typu kontrolních činností by měla organizace zvážit vynaložené náklady a efekt, které tyto kontroly přinesou. Mezi pět základních kontrolních činností poté patří *fyzická kontrola majetku*, *segregace povinností*, *systém autorizace*, *nezávislé kontroly* a *kontrola dokumentace a záznamů* (Albrecht, 2011, s. 42-115). S ohledem na preventivní opatření je vhodné zavést segregaci povinností například do oblasti nákupů, příjmů a mezd. Zavedení omezeného přístupu se pak doporučuje v účetním softwaru (Wiersema, 2015, s. 30-31). Fyzické kontroly představují omezený přístup k zařízením, mající podobu například v uzamykatelných zásuvkách nebo trezorech (Trenerry, 1999, s. 28). Nezávislé kontroly a kontroly dokumentů se ovšem řadí již mezi kontroly detektivní (Albrecht, 2011, s. 114). Tyto kontroly tedy k výskytu podvodů nebrání, ale poukáží na existenci klamání. Kombinaci preventivních a detektivních kontrol je ovšem vhodné používat, neboť tento postup zvýší efektivitu řízení rizik (Dubis, 2009, s. 21).

Aby bylo vedení schopné posoudit efektivnost a účinnost operací, je nutné mu zajistit spolehlivé informace. Obsah zpráv je tvořen provozními a finančními operacemi, jejichž zdrojem je převážně účetní systém. K tomu, aby byly informace spolehlivé, je nutné, aby byl účetní systém kvalitní. Takovýto systém by tak měl identifikovat všechny důležité transakce organizace, měl by zachytit důležitá data a ty pak zaznamenávat a zpracovávat klasifikací, sumarizací a agregací (Turner, 2009, s. 99). Zavedením kvalitního účetního systému je i zde důležitým prvkem k potlačení faktoru příležitosti, neboť podvody jsou nejčastěji ukryty v transakčních záznamech a je tedy pravděpodobné, že na základě dobrého účetního systému budou tyto podvody snáze odhaleny z důvodu přítomnosti auditní stopy (Albrecht, 2011, s. 42).

Poslední položku představuje proces, který celý systém kontroly monitoruje. Monitorovací činnosti jsou nutné k celkovému zhodnocení efektivity a k detekci nedostatků, které by se

vedení mělo pokusit vylepšit. Z důvodu neustálého zhodnocení interní kontroly je efektivnější zavést spíše proces sledování, který by měl být průběžný a pravidelný (Trenerry, 1999, s. 14).

Jak je možné si z výše uvedených kroků povšimnout, strategie prevence vnitřní kontroly se opírá o vytváření podmínek, které mohou organizaci k opatřením proti podvodům nasměrovat nebo naopak podvody odhalit v počátečních fázích před vznikem ztrát (Yuniarti, 2017, s. 114). S ohledem na určité omezení, které vnitřní kontrola může zaznamenat ze strany vedení, zaměstnanců nebo vznikem tajných dohod, se systém kontroly stává nutným, avšak ne zcela dostatečným prvkem řízení (Dimitrijevic, 2015a, s. 36-37). Z tohoto důvodu je tedy vhodné vnitřní kontrolu kombinovat s některými dalšími preventivními strategiemi. S ohledem na potlačení příležitosti k podvodu je vhodné se zaměřit na efektivní kontrolní prostředí, kontrolní postupy a efektivní účetní systém. Pokud systém kontroly tyto tři požadavky splňuje, existuje dostatečná pravděpodobnost, že cílů organizace bude dosaženo a výskyt podvodů bude omezen (Albrecht, 2011, s. 114).

2.1.4. Etický kodex

Mimo preventivní techniky a systém vnitřní kontroly je důležité, aby společnosti, které chtějí podvodům předcházet, vlastnily etický kodex. Etický kodex by se měl vztahovat na všechny zaměstnance, měl by řešit důvěryhodnost informací, střet zájmů nebo prevenci a odhalování podvodů (Moore, 2010, s. 73). Obsah kodexu je tak tvořen sadou dokumentů, jeho součástí jsou etické pokyny a za jejich vystavení a prosazování je odpovědné vedení (Turner, 2009, s. 80). Implementací těchto zásad je vedoucím pracovníkům dán nejlepší způsob, jak informovat zaměstnance o tom, co je a není ve společnosti přijatelné (Lendez, 1999, s. 54).

Při tvorbě etického kodexu by poté neměly být opomíjeny rizikové oblasti společnosti a s tím i spojené hodnoty a chování potřebné k dodržení předpisů a zákonů. Kodex by měl být psán srozumitelnou formou a měl by zvážit i hodnoty, které propojují organizaci se zainteresovanými stranami (Ferrell, 2012, s. 224).

Důležitost vlastnictví etického kodexu zmiňuje také americký zákon *Sarbanes-Oxley Act*, dle kterého musí od roku 2002 vlastnit kodex chování všechny veřejné obchodní společnosti působící na amerických burzách pod jurisdikcí SEC. Cílem tohoto zákona je tak podpořit zaměstnance v nejednoznačných situacích a formulovat pravidla pro všechny zúčastněné strany (Wulf, 2011, s. 13).

2.2. Detekce finančních podvodů

Smyslem detektivních strategií je zaznamenat určité důkazy nebo i případná varování, že k výskytu podvodů buď dochází nebo k nim již došlo. Detekční metody tak musí být převážně flexibilní, adaptabilní a musí se průběžně měnit, aby mohly reagovat na rizikové změny prostředí (Dubis, 2009, s. 21). Mnoho detekčních metod poté souvisí i s preventivními technikami, neboť se některé detekční kroky s preventivními překrývají. Ačkoliv jsou preventivní kroky méně nákladné a efektivnější, ne vždy podvodu zabrání, a ne vždy je prevence v organizaci možná, proto jsou strategie detekce pro společnost nezbytný prvkem (Goldmann, 2009, s. 119).

Způsobů, které dokáží podvody odhalit nebo odhalit jejich potenciální výskyt, existuje mnoho. V rámci jednoduchosti budou detekční přístupy rozděleny do dvou podkapitol. V první podkapitole budou zmíněny základní manuální přístupy, ve druhé pak techniky pokročilé, postavené na modelech umělé inteligence.

2.2.1. Základní přístupy detekce

Mezi základní techniky, které vedení organizace může využít, jsou tradiční techniky odhalování, konané prostřednictvím webových formulářů, mailů nebo rozhovorů. Tyto způsoby mohou zahrnovat způsoby detekce, kdy zaměstnanec v rámci odsouhlasení etického kodexu obdrží povinnost jakýkoliv výskyt podvodného jednání hlásit. Jiný způsob, jak podvod odhalit, může být dán i prostřednictvím horké linky nebo vedením výstupního pohovoru. Osoba ukončující pracovní poměr může pomoci identifikovat výskyt podvodu nebo poskytnout informace ohledně podmínek, které k podvodům přispívají (Dubis, 2009, s. 21).

Mimo hlášení zaměstnanců, šetření ze strany vedení nebo detekce, která může být i zcela náhodná, může být účinným nástrojem detekce i interní audit (Seetharaman, 2004, s. 1064). Role interního auditu ve spojení s detekcí podvodů je v organizaci možná prostřednictvím informace, kterou audit organizaci poskytuje. Ačkoliv interní audit v rámci vnitřní kontroly posuzuje rizikové oblasti vnitřního kontrolního prostředí, může být zapojen i do procesu spolehlivého finančního výkaznictví (Kranacher, 2011, s. 46). Primárním cílem interního auditu není tedy podvody odhalit, ale auditoři mohou při provádění běžné kontroly detekovat v účetních uzávěrkách jisté nesprávnosti nebo opomenutí. Na tomto místě je ovšem důležité, aby auditoři disponovali i patřičnými znalostmi o ukazatelích podvodu, znali expozici podvodu ve specifických termínech nebo začlenili do auditu kroky, které příznaky podvodu pravděpodobně odhalí (Seetharaman, 2004, s. 1064).

Oproti internímu auditu, který má tendenci být spíše preventivním a kontrolním prostředkem, je k detekci zaměřen spíše audit externí (Vallabhaneni, 2005, s. 441). Externí audity druhých a třetích stran mohou být k organizaci buď v bezprostředním vztahu nebo mohou být prováděny jednotlivci z odlišné organizace (Hollenback, 2007, s. 12). Typ tohoto auditu je poté zaměřen na kontrolu účetních uzávěrek a auditoři zde mohou podvody detekovat prostřednictvím analytických postupů v podobě vývojových diagramů, statistického odběru vzorků nebo analýzy finančního výkaznictví (Vallabhaneni, 2005, s. 441).

S ohledem na audit je na tomto místě vhodné zmínit i možnost využití auditu, který není plánovaný, a to zejména v oblastech s vysokým rizikem výskytu podvodu (Dubis, 2009, s. 22). Typ tohoto auditu probíhá bez větších upozornění a jeho výhoda k možnostem detekce je dána nízkou připraveností ze strany pachatelů, kteří nemají dostatek času k zakrytí nedostatků (Hollenback, 2007, s. 13).

Ačkoliv je auditorská činnost z pohledu investorů činností záslužnou, neboť díky tomuto konání investoři získávají ujištění ve správnosti finančních výkazů, mnohdy tento způsob detekce není dostačujícím způsobem, jak podvody v organizacích odhalit. Auditoři nemusí mít dostatek znalostí k odhalení podvodu (Huang, 2016, s. 4-5), případně může být zaznamenán i konflikt zájmů mezi vlastním zájmem auditora a jeho profesní odpovědností (Moore, 2006, s. 10). Celková auditorská činnost tak může být v důsledku těchto jevů považována za ne zcela dostatečným prvkem detekce (Huang, 2016, s. 5).

Z výše uvedených přístupů si je tak možné povšimnout, že důležitým prvkem v detekci jsou ukazatelé podvodu, tedy „červené vlajky“ (*red flags*), dle kterých mohou auditoři, management nebo i zaměstnanci případná podvodná jednání detekovat. Zde nutno poznamenat, že výskyt ukazatelů nutně neznamená výskyt podvodného jednání. Po fázi detekce je potřebné shromáždit potřebné důkazy, které dosvědčují, zda je příčinou symptomu skutečně podvod nebo jiný faktor, který podvodem být nemusí. Mezi základní kategorie těchto symptomů patří *analytické anomálie*, vyjadřující množství nebo poměry, jejichž vykazování je nějakým způsobem neobvyklé, *účetní anomálie*, vyplývající z procesů nebo postupů v účetním systému, *slabá vnitřní kontrola*, *neobvyklé chování*, *změna životního stylu* a *tipy a stížnosti* (Carmichael, 2007, s. 63; Albrecht, 2011, s. 214).

2.2.2. Pokročilé přístupy detekce

Oproti tradičním způsobům manuální detekce, které jak bylo zmíněno v předchozí kapitole, mohou zahrnovat detekční techniky nejen nepřesné ale i nákladné, umožňují pokročilé přístupy

především praktičtější přístup k detekci nad větším množstvím dat a menší časovou náročnost (West, 2016, s. 47).

Mezi tyto přístupy lze tak zařadit techniky pro správu podvodů z oblastí umělé inteligence, dolování dat, expertních systémů nebo technik strojového učení (Girish, 2002, s. 49). S přístupem odhalení podvodů pomocí těchto metod lze tak mimo zajištění vysoké přesnosti detekce zajistit i minimální potřebu zásahu člověka a zároveň zajistit schopnost vlastního učení a přizpůsobení detekčního systému měnícím se vzorům podvodu (Alford, 2013, s. 14). Mimo výše uvedené výhody je tento přístup také vhodným nástrojem pro vládní regulátory, auditorské společnosti nebo investory, neboť je tímto přístupem subjektům poskytnuta lepší orientace v určení společností s vyšším rizikem podvodu (Cecchini, 2010, s. 1158).

K rozsáhlému přehledu využití těchto pokročilých detekčních přístupů se poté věnují autoři West (2016, s. 47-66) a Ngai (2011, s. 559-569), kteří rekapitulují komplexní přehledy studií detekce prostřednictvím metod z oblasti dolování dat. Oba autoři rozlišují detekci dle typu klasifikace finančních podvodů, přičemž s ohledem na podvody konané v rámci společností West (2016, s. 51) shrnuje jako vhodné metodiky k detekci podvodů ve finančních výkazech metody rozhodovacích stromů, metodu podpůrných vektorových strojů, neuronové sítě, genetické algoritmy nebo *text mining*. K detekci podvodů vyskytujících se v cenných papírech a komoditách pak bayesovskou síť nebo *process mining*.

3. KOMISE PRO CENNÉ PAPÍRY A BURZY

S ohledem na experimenty provedené v další části této práce, bude následující kapitola věnována nezávislé vládní agentuře SEC, působící na amerických kapitálových trzích a o jejím přehledu vynucovacích opatření, převážně zaměřených proti veřejným obchodním společnostem z období let 2010-2015. První část kapitoly bude věnována krátkému popisu SEC. Druhá část kapitoly bude zaměřena na vybrané ukázky z přehledu vynucovacích opatření, poskytnutých empirickou databází pro vymáhání práv z oblastí cenných papírů (SEED), evidující informace o vynucovacích akcích, které SEC proti vybraným subjektům v průběhu let vedla.

3.1. Úvodem o SEC

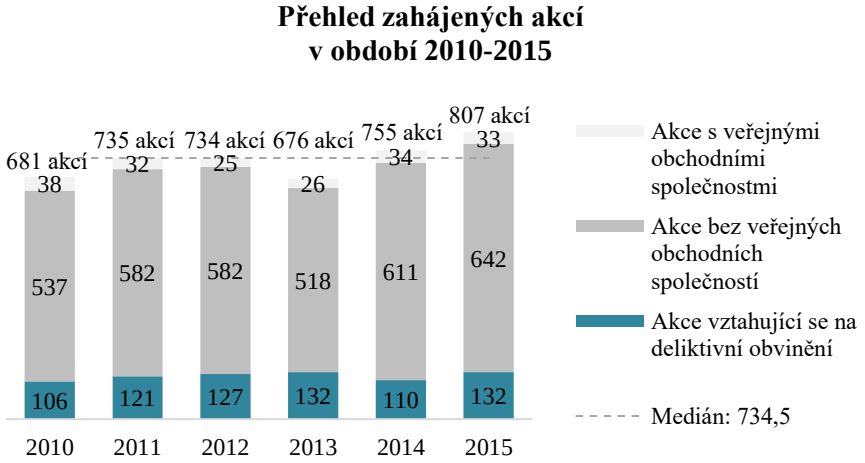
Komise pro cenné papíry a burzy (SEC) představuje americkou vládní agenturu, jejíž vznik je připisován k roku 1934, kdy tato komise vzešla z vydání zákonu o burzách cenných papírů (*Securities Exchange Act of 1934*). Prostřednictvím tohoto zákona tak bylo SEC poskytnuto vymáhací právo pro dodržování legislativy o cenných papírech z roku 1933 (*Securities Act of 1933*), které stanovuje požadavky na zveřejnění důležitých informací o cenných papírech, poskytnutých k veřejnému prodeji. Dle zákonu o cenných papírech (1933) jsou tak společnosti povinné před nabídkou cenných papírů veřejnosti projít registračním procesem, ve kterém jsou ze strany SEC kontrolovány korektnosti, informující investory o vlastnostech společnosti, typu cenných papírů a základním řízení společnosti. Odpovědnost za registrační formuláře nese registrující společnost, vztahující se na emitenta (Welytok, 2006, s. 41-42).

Prostřednictvím zákona o burze cenných papírů (1934) tak získává SEC právo registrace a držení dohledu nad makléřskými společnostmi, burzou cenných papírů nebo zúčtovacími agenturami. Mimo požadavků na zveřejnění významných informací, zajišťujícím investorům lepší proces rozhodování, zákon o burzách vymezuje i určité druhy jednání, které jsou na trhu s cennými papíry nepřípustné. Z těchto důvodu je tak Komisi SEC (dále jen Komise) udělena i pravomoc vést vůči vinným subjektům nebo konkrétním jednotlivcům právní opatření (Welytok, 2006, s. 41-42).

3.2. Ukázky z přehledu vynucovacích akcí

Grafické znázornění (Obrázek 7) poukazuje na celkový počet vynucovacích akcí, které Komise v období let 2010-2015, proti veškerým subjektům vedla. Jak je možné vidět, v průběhu sledovaných let počet vynucovacích opatření průběžně rostl. Výjimkou se stal rok

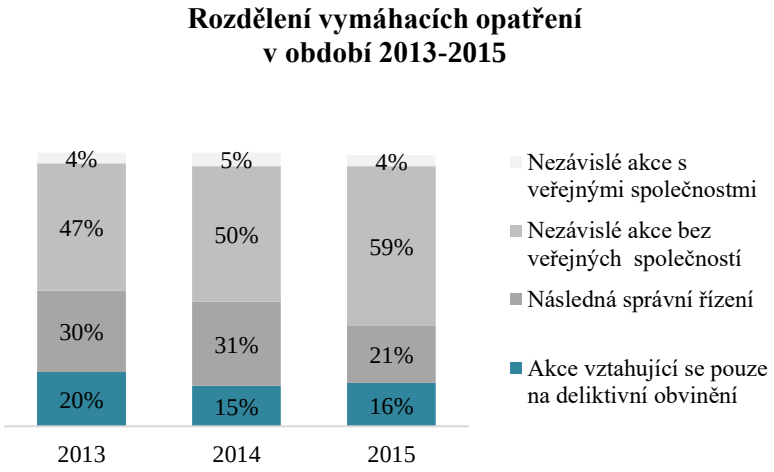
2013, kde bylo provedeno, oproti předchozím rokům, vynucovacích opatření méně. Průměrný počet vymáhacích akcí se v průběhu let 2010-2015 pohyboval přibližně okolo 735 akcí ročně, přičemž nejvýraznější nárůst byl zaznamenán v průběhu let 2013-2015, kde byl navýšen počet vynucovacích akcí až na 807 případů. Rok 2015 se tak stal nejvíce vytíženým rokem Komise ze všech předchozích sledovaných let.



Obrázek 7: Přehled vynucovacích akcí v období let 2010-2015

Zdroj: upraveno dle (Choi, 2016, s. 3)

Příčina zmíněného vzrůstu vynucovacích akcí v letech 2013-2015 byla dána důsledkem zvýšeného počtu vykonaných nezávislých opatření ve věci porušení zákona o cenných papírech (Choi, 2016, s. 4). Tento nárůst tvořil v roce 2015 až 63 % všech celkových opatření. Ve sledovaném horizontu tří let procentuální počet následných správních řízení, vedených dodatečně vůči jednotlivcům, poklesl o 9 %. Procentuální podíl vymáhacích opatření proti veřejným obchodním společnostem zůstal v tomto směru relativně konstantní (Obrázek 8).



Obrázek 8: Rozdělení vymáhacích opatření v období 2013-2015

Zdroj: upraveno dle (Choi, 2016, s. 4)

V následujícím textu již budou zohledněny pouze akce týkající se obžalovaných veřejných společností. V úvahu zde tedy nebudou brány obvinění vztahující se k neveřejným dceřiným společnostem a akce vedené proti jednotlivcům.

Níže uvedené vyjádření (Tabulka 1) poukazuje na kategorie obvinění, které Komise proti veřejným společnostem v průběhu let vedla. Nejčastějším typem tvrzení byl výrok vedený ve věci porušení emitentního vykazování informací, dané zákonem o cenných papírech. Největší nárůst tohoto typu pochybení byl poté zaznamenán k roku 2013 a dále, neboť důvodem tohoto zlomu bylo znovuzavedení iniciativy Komise zaměřené na zlepšení korektnosti vykazování informací v účetních uzávěrkách (Choi, 2016, s. 12).

Počet obvinění porušující zahraniční zákon o korupčních praktikách se v průběhu sledovaných let udržoval relativně stabilně pod 50% hranicí. Pouze v roce 2011 byl zaznamenán větší nárůst toto typu tvrzení, které převládlo nad typem žalob vedených proti emitentovi. Porušení zahraničního zákonu o korupčních praktikách pak spolu s porušením emitentního vykazování informací tvořil v průměru až 90 % všech obvinění za období let 2010-2014. Samostatně stojící obvinění porušující zákon o korupčních praktikách poté dosahovalo v roce 2015 stejného podílu jako v celkovém průměru za období let předchozích.

Dalšími, již méně se vyskytujícími typy obvinění, uvádí Komise tvrzení vedené ve věcech nezákonných manipulací s tržními cenami cenných papírů, prodeje neregistrovaných cenných papírů, porušení odpovědnosti makléřských společností vůči zákazníkovi nebo odcizení zákaznických fondů a cenných papírů (*What We Do*, 2013).

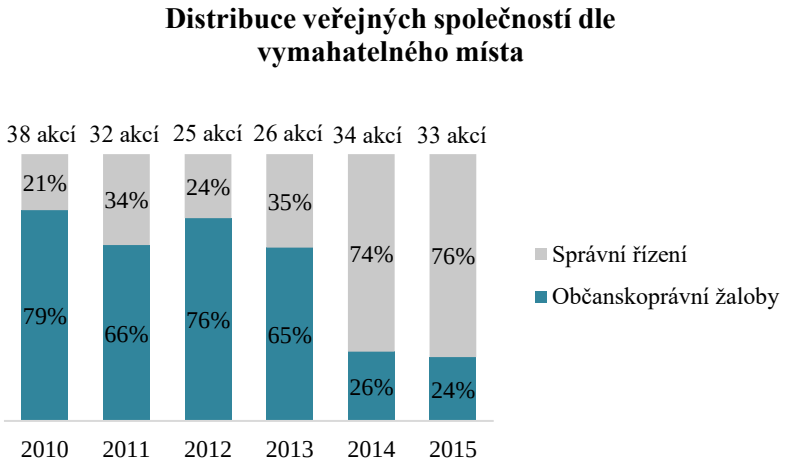
Tabulka 1: Typy tvrzení vedených proti veřejným obchodním společnostem

Typ tvrzení	Průměr 2010–2014	2010	2011	2012	2013	2014	2015
Emitentní vykazování informací	57 %	47 %	44 %	48 %	73 %	74 %	52 %
Zahraniční zákon o korupčních praktikách	33 %	32 %	53 %	40 %	19 %	21 %	33 %
Jiné	2 %	5 %	0 %	4 %	0 %	0 %	6 %
Investiční poradce/investiční společnosti	2 %	11 %	0 %	0 %	0 %	0 %	3 %
Manipulace na trhu	1 %	0 %	3 %	4 %	0 %	0 %	3 %
Makléř	1 %	0 %	0 %	0 %	4 %	3 %	3 %
Nabídka cenných papírů	3 %	5 %	0 %	4 %	4 %	0 %	0 %
Cenné papíry/veřejné penzijní fondy	1 %	0 %	0 %	0 %	0 %	3 %	0 %
Insider Trading	0 %	0 %	0 %	0 %	0 %	0 %	0 %
Počet akcí	31	38	32	25	26	34	33

Legenda	0 %	1-10 %	11-20 %	21-50 %	51-100 %

Zdroj: upraveno dle (Choi, 2016, s. 6)

Níže uvedené grafické znázornění (Obrázek 9) poukazuje na rozlišení veřejných společností dle prostředí, kde bylo konečné vynucovací právo provedeno. Jak Komise uvádí, důvodem k rozhodnutí předat obvinění federálnímu soudu nebo jej řešit ve formě vlastního správního řízení může být dáno typem sankce nebo úlevy, o kterou se žádá (*What We Do*, 2013).



Obrázek 9: Distribuce veřejných společností dle vymahatelného místa

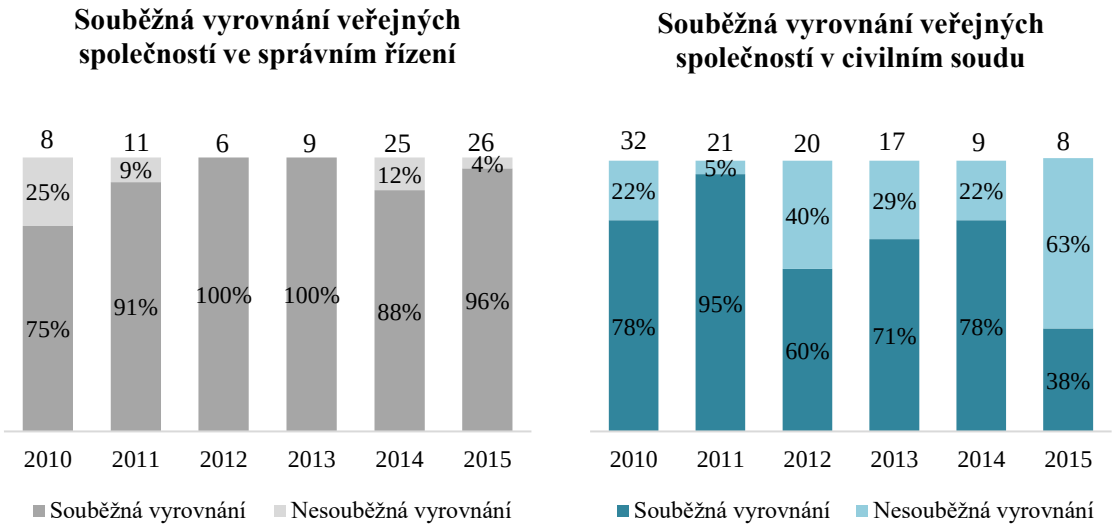
Zdroj: upraveno dle (Choi, 2016, s. 6)

Je možné si povšimnout, že ve sledovaných letech nebylo místo výkonu rozhodnutí zcela jednoznačné, avšak stále více bylo Komisí prosazované správní řízení. Do roku 2013 Komise předložila více než 65 % případů k občanskému soudu. Od roku 2014 počet civilních žalob výrazně klesl a v posledních dvou uvedených letech se pro veřejné společnosti stalo rozhodovacím místem vymáhání správní řízení, jehož nárůst byl od počátku sledovaného roku takřka trojnásobný.

K vysvětlení příčiny neustále se zvětšujícího prosazování vlastního správního řízení Komise před podáním žalob k federálnímu soudu může pomoci grafické znázornění uvedené dále (Obrázek 10). Níže uvedené grafy vyjadřují procentuální srovnání souběžných vyrovnání, tedy procentuální počet podaných žalob, které byly vyřešeny v tentýž den, kdy byly zahájeny. Ze strany Komise je tedy znatelné, že jím vedené správní řízení dosahuje v tomto směru průměrné úspěšnosti přes 90 %. Naproti tomu obvinění podané před federální soud dosahuje průměrné úspěšnosti pouze necelých 70 %. Na tomto místě je tedy znatelný důvod rozšířeného využívání správního řízení ze strany Komise.

Z detailnějšího popisu souběžných vyrovnání si je možné povšimnout i jisté shody v roce 2010 u obou alternativ vymahatelných míst. V tomto roce byl podíl žalovaných veřejných společností se souběžným urovnáním přibližně stejný, a to jak u občanskoprávních, tak

správních řízení. Procentuální podíl občanskoprávních žalob oproti správnímu řízení k roku 2015 výrazně poklesl, a to až k 38 %. Čísla uvedená nad jednotlivými sloupci představují počet žalovaných společností.



Obrázek 10: Souběžná vyrovnání žalob ve správním řízení a civilním soudu

Zdroj: upraveno dle (Choi, 2016, s. 8)

4. STROJOVÉ UČENÍ

Obsah následující kapitoly bude věnován popisu metod použitých v další části práce, přičemž tato kapitola bude nejdříve zaměřena na základní koncepci strojového učení. Další části již budou věnovány popisu konkrétních metod, zahrnující modely neuronových sítí, podpůrných vektorových strojů a rozhodovacích stromů. Tyto tři metody budou dále využity i v rámci meta učení, a sice v podobě metod *Bagging*, *Boosting* a *Stacking*, jejichž popis bude rozebrán ve druhé části kapitoly.

4.1. Koncept strojového učení

Obecnou myšlenku strojového učení, vedené jako podoblast umělé inteligence, lze prezentovat schopností počítače učit se určitou úlohu z trénovacích vzorů, díky čemuž je systému dána možnost řešit tentýž typ úlohy i s novými daty (Louridas, 2016, s. 110). V podstatě tak na tomto místě dochází k programování počítačů, které jsou díky učení se z příkladů a zkušeností z minulosti, schopny reagovat na změny, čímž dochází i ke zlepšení jejich výkonu (Alpaydin, 2010, s. 3; Geetha, 2016, s. 147).

Výstižně koncept komentuje Mitchell (1997, s. 2), jenž uvádí: „*Stroj se učí ze zkušenosti Z s ohledem na určitou třídu úkolů U při výkonnostní míře V, pokud se jeho výkonnost na úkolech U, měřená pomocí V, zlepšuje se zkušenostmi Z*“.

Užitečnost tohoto přístupu je možné uplatnit v mnoha oblastech, počínaje rozpoznáváním řeči, obrazu nebo robotiky, avšak stimul k využití tohoto konceptu může být dán schopností systému přizpůsobit se právě díky učení změnám prostředí, díky čemuž nemusejí být návrhářem systému předvídány postupy a řešení pokrývající všechny situace. Jiná příčina může vyplývat i z postradatelnosti vhodného algoritmu pro realizaci určitých druhů úloh (Alpaydin, 2010, s. 1-3) nebo z možnosti analyzovat rozsáhlé datové soubory, které jsou pro manuální rozbor příliš složité (Geetha, 2016, s. 147).

S ohledem na strukturu trénovacích vzorů a požadovaný výstup, lze algoritmy strojového učení dělit do dvou základních kategorií, a sice na typ *učení s učitelem* a *učení bez učitele* (Dua, 2011, s. 7). Vzhledem k tomu, že data v této práci budou označena výstupními třídami, bude v nadcházejícím textu věnován větší prostor učení s učitelem.

Učení s učitelem vyjadřuje formu učení, kdy je systému poskytnut soubor trénovacích vzorů $\mathcal{D}$ ve formě vstupních $\boldsymbol{x}_i \in X$ a požadovaných výstupních $y_i \in Y$ vztahů, $i = 1, 2, \ldots, N$, kde N vyjadřuje velikost trénovací sady, přičemž cílem této formy učení je naučit systém mapování

ze vstupů x na výstupy y. Jednotlivé vstupy $\boldsymbol{x}_i$ pak obvykle tvoří d-dimenzionální vektor hodnot, obecně představující komplexní, strukturovaný objekt (Murphy, 2012, s. 2):

$$\mathcal{D} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{N} \tag{4.1}$$

Dále je v rámci tohoto učení rozlišováno mezi dvěma typy úloh, a sice mezi úlohou *klasifikační*, kdy $y_i \in \{1, \dots, C\}$, kde C vyjadřuje počet tříd a úlohou *regresní*, kde $y_i \in \mathbb{R}$. V případě klasifikační úlohy pak může být rozlišováno mezi klasifikací binární, jestliže $C = 2$, diskrétní, kdy $C > 2$ nebo klasifikací vícenásobnou, jestliže se třídy mezi sebou vzájemně nevylučují (Murphy, 2012, s. 2-3).

Formálněji lze poté učení s učitelem vyjádřit přístupem, kde je úkolem vzhledem k trénovací sadě $(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \dots, (\boldsymbol{x}_N, y_N)$, kde $y_i = f(\boldsymbol{x}_i)$, nalézt takovou hypotézu $h \in H$, která by se skutečné funkci f co nejvíce přiblížila. Proces učení tak vyjadřuje hledání vhodné hypotézy $h: X \to Y$, jež bude přesná i na datech mimo trénovací vzory (Russell, 2010, s. 695-696; Bunt, 2004, s. 28).

Teoreticky tak lze na tuto problematiku nahlížet jako na optimalizační problém minimalizace chyby funkce, kde touto nejvíce rozšířenou minimalizovanou funkcí je *chyba empirická* (Madani, 2011, s. 363). Empirická chyba L vyjadřuje poměr trénovacích vzorů, kdy předpověď hypotézy neodpovídá požadovaným hodnotám (Alpaydin, 2010, s. 24), přičemž její hodnotu lze vyjádřit rozdílem mezi odhadem systému $h(\boldsymbol{x}_i)$ a y_i (Russell, 2010, s. 711).

Na základě výše uvedených poznatků lze tak říci, že vzhledem k trénovacím vzorům a prostoru hypotéz H je cílem učení nalézt takový algoritmus s výpočtem h, pro který bude empirická chyba minimální (Bunt, 2004, s. 29):

$$\hat{h} = \arg\min_{h \in H} \frac{1}{N} \sum_{i=1}^{N} L(y_i, h(\boldsymbol{x}_i)) \tag{4.2}$$

Vhodné modely pro klasifikační typ úloh zahrnují neuronové sítě, klasifikační stromy nebo logistickou regresi. V rámci regresní úlohy lze využít modely lineární regrese, bayesovské sítě nebo regresní stromy (Louridas, 2016, s. 113).

Trénovací vzor $\mathcal{D}$ v případě *učení bez učitele* na rozdíl od předchozího typu učení známé odpovědi neposkytuje, a tudíž zde dochází k postradatelnosti informace o chybovém signálu. Cílem tohoto učení je tak nalézt ve vstupních datech podobnosti (Murphy, 2012, s. 2; Geetha, 2016, s. 148) tak, aby vstupy, které mají něco společného, mohly být rozděleny do tříd (Marsland, 2015, s. 6):

$$\mathcal{D} = \{(\boldsymbol{x}_i)\}_{i=1}^{N} \tag{4.3}$$

Louridas (2016, s. 113) rozděluje modely učení bez učitele do dvou směrů, a sice na modely týkající se shlukování a modely spojené s redukcí dimenze. K úloze shlukování lze využít metodu k-průměrů, hierarchické shlukování nebo genetické algoritmy. Ke snížení dimenze datové sady pak například analýzu hlavních komponent nebo neuronové sítě.

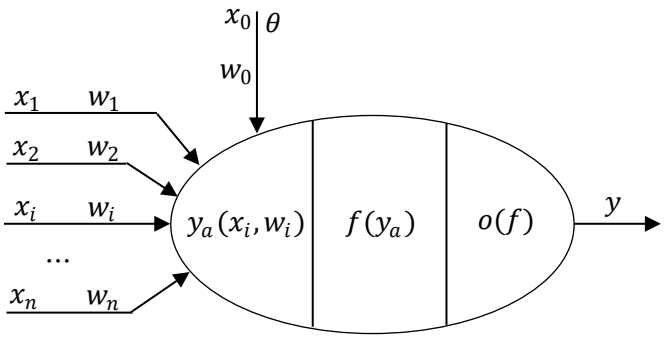
Autoři (Schuld, 2015, s. 3; Marsland, 2015, s. 6; Geetha, 2016, s. 148) uvádí k výše uvedeným dvěma základním typům i další typ učení, a sice *učení posilované*. Tento případ využívá k učení signály ve formě odměn a trestů (Murphy, 2012, s. 2). Agent tvořící rozhodnutí zde spolupracuje s prostředím, ve kterém musí řešit problém, přičemž v rámci jím zvolených strategií vedoucích k vyřešení otázky, získá odměnu nebo trest. Po několika provedených pokusech je tak zjištěn optimální sled akcí, které celkovou odměnu maximalizují (Alpaydin, 2010, s. 447).

4.2. Základní algoritmy učení

4.2.1. Umělé neuronové sítě

Model umělé neuronové sítě představuje výpočetní strukturu napodobující chování lidského mozku, využívající ke grafické prezentaci synapsí a neuronů soubor propojených vrcholů (Ngai, 2011, s. 563). Mezi hlavní výhody neuronových sítí, včetně schopnosti adaptace, generalizace a schopnosti učit se, lze zařadit odolnost vůči chybám (Jain, 1996, s. 31).

V souvislosti s finančními podvody jsou neuronové sítě především využívány k detekci podvodů v rámci korporátní oblasti, v oblastech pojišťovacích podvodů a podvodů s kreditními kartami (Ngai, 2011, s. 563).



Obrázek 11: Matematický model neuronu

Zdroj: (Olej, 2010, s. 50)

Na výše uvedeném znázornění (Obrázek 11) je uveden primární prvek neuronové sítě v podobě formálního modelu neuronu kde:

- x_i – vstupy neuronu (výstupy z předchozí vrstvy), $i = 1,2, \dots, n$,
- n – počet vstupů (počet neuronů z předchozí vrstvy),
- w_i – synaptické váhy,
- y_a – vstupní potenciál neuronu,
- f – aktivační funkce neuronu,
- o – výstupní funkce,
- θ – práh neuronu,
- y – výstup neuronu.

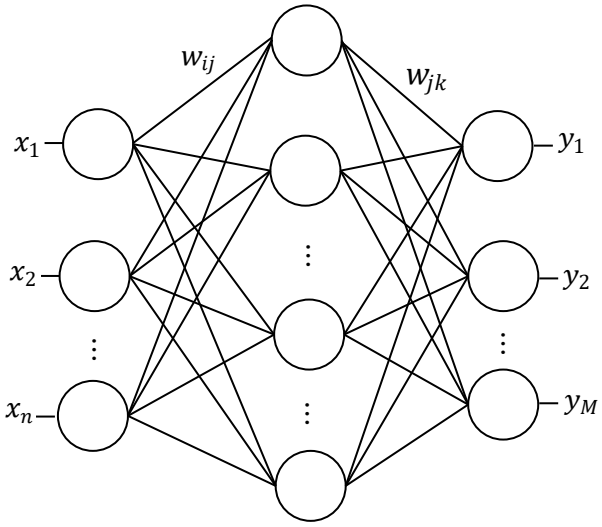
K jednotlivým vstupům $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$ jsou přiřazeny synaptické váhy $\boldsymbol{w} = (w_1, w_2, \dots, w_n)$, udávající informaci o důležitosti jednotlivých vstupů na výstup z neuronu. Vstupní potenciál y_a vzniká výsledkem váženého součtu vstupů $\boldsymbol{x}_i$ a příslušných vah $\boldsymbol{w}_i$, kde je následně výsledek vzniklého potenciálu porovnán s hodnotou prahu θ, určující stav aktivity neuronu. Prostřednictvím aktivační funkce f, omezující rozsah výstupního signálu na konečnou hodnotu, je poté výsledek ze vstupního potenciálu odvozen na výstupní hodnotu neuronu (Olej, 2010, s. 51-52). Konečnou hodnotu výstupu formálního neuronu lze vyjádřit následujícím způsobem (Russell, 2010, s. 728):

$$y = f(y_a) = f\left(\sum_{i=0}^{n} x_i w_i + \theta\right) \tag{4.4}$$

Aktivační funkce může mít několik podob a její výběr je odvozen od typu řešené úlohy nebo umístění neuronu v síti. Mezi možné případy této funkce lze zařadit funkci skokovou, funkci sigmoidální nebo gaussovu (Olej, 2010, s. 52). V případě aktivační funkce mající podobu ostré prahové hodnoty je jednotka neuronu nazývána *perceptronem*, v případě logistické funkce lze použít název *sigmoid perceptron* (Russell, 2010, s. 729).

Vícevrstvá perceptronová síť (Obrázek 12) poté představuje typ dopředné neuronové sítě s možností aplikace učení na úlohy klasifikačního i regresního charakteru (Alpaydin, 2010, s. 232). Jednotlivé neurony jsou uspořádány do vrstev, kde vstup do vstupních jednotek (neuronů) tvoří informace z datového souboru a mezilehlé neurony tvoří nejméně jednu skrytou vrstvu. Výstupní vrstva pak poskytuje odpovědi pro danou sadu vstupních hodnot (Marini, 2007, s. 119).

Haykin (1999, s. 157) na tomto místě jmenuje tři charakteristické vlastnosti vícevrstvé perceptronové sítě, a sice kdy (1) model každého neuronu zahrnuje nelineární aktivační funkci, v obvyklé podobě sigmoidální nelinearity, (2) vrstva skrytých neuronů, umožňující síti učit se nelineární úlohy, není součástí vstupu ani výstupu a (3) vysoký stupeň synaptických spojů.



Obrázek 12: Vícevrstvá perceptronová síť s jednou skrytou vrstvou

Zdroj: upraveno dle (Olej, 2010, s. 58)

Nejčastější přístup k učení je zde definován prostřednictvím algoritmu zpětného šíření chyby (Mitchell, 1997, s. 83), kde je cílem iterativně měnit synaptické váhy mezi neurony ve směru minimalizace chyby E, definované jako mocnina rozdílu mezi požadovaným a aktuálním výsledkem z výstupních neuronů nasčítané v průběhu učení (Marini, 2007, s. 119). Následující vyjádření představuje rovnici pro adaptaci vah (Mařík, 2003, s. 210):

$$\Delta w_{ij}(t) = -\eta \frac{\partial E}{\partial w_{ij}} + \alpha \Delta w_{ij}(t-1) \tag{4.5}$$

Změna synaptických vah v rámci t-té epochy závisí na parciální derivaci celkové chyby podle této váhy a konstanty η, vyjadřující rychlost učení a změně téže váhy během předchozí iterace prostřednictvím momenta α. Pro výpočet parciální derivace chyby E podle vah mezi neurony ve skryté vrstvě, je potřebné využít řetízkového pravidla derivace (Marini, 2007, s. 119). Bližší rozbor této problematiky uvádí Haykin (1999, s. 161) nebo Mařík (2003, s. 209).

Každý skrytý nebo výstupní neuron vícevrstvého perceptronu je tedy navržen tak, aby konal dva typy výpočtů, a sice výpočet *funkčního signálu*, který se objevuje na výstupu neuronu a výpočet *odhadu vektoru gradientu*, který je potřebný pro zpětný průchod sítí (Haykin, 1999,

s. 160). Mezi charakteristiky ovlivňující chování sítě lze zařadit velikost rychlosti učení, momentum, šum a počet neuronů (Mařík, 2003, s. 210).

4.2.2. Rozhodovací stromy

Rozhodovací stromy představují hierarchickou datovou strukturu s možností aplikace na typ klasifikačních a regresních úloh (Alpaydin, 2010, s. 185). Výhoda této metody je dána především v jednoduchosti implementace a přehlednosti výsledků. Učení zde vyžaduje nízkou výpočetní náročnost, přičemž použití této metody k odhalení finančních podvodů se zdá být vhodným prostředkem vzhledem k jejímu úspěšnému využití při řešení detekce podvodů spojené s finančními výkazy a kreditními kartami (West, 2016, s. 52-,58).

Formálněji pak rozhodovací strom představuje neorientovaný graf, kde jsou každé dva uzly propojeny právě jednou hranou. Výchozí uzel bez příchozích hran je nazýván kořenem a koncový uzel s příchozí, ale ne odchozí hranou, je nazýván listem. Každý uzel mimo konečných listů pak obsahuje rozhodovací funkci (Schuld, 2015, s. 12), kdy je při zavedení vstupu v každém takovém uzlu proveden test, jehož důsledkem v závislosti na výsledku je provedení jedné z větví stromu. Tento proces začíná u kořene a rekurzivně se opakuje, dokud není dosáhnuto listu (Alpaydin, 2010, s. 185-186). Kromě cíle nalezení optimálního stromu prostřednictvím minimalizace chyby lze zohlednit i další účelové funkce v podobě minimalizace počtu uzlů stromu nebo minimalizace průměrné hloubky (Rokach, 2005, s. 477-478).

Obvyklé algoritmy (ID3, C4.5, CART) postupují při konstrukci stromu dle základního algoritmu TDIDT (Fisk, 2015, s. 86), kdy je v každé iteraci rozdělení trénovacích vzorů výsledkem diskrétní funkce vstupních atributů. Každý uzel tak dělí trénovací množinu na menší podmnožiny, dokud není rozdělení dostatečné nebo není splněné kritérium zastavení (Rokach, 2005, s. 478). Berka (2003, s. 86) definuje proces konstrukce stromu následovně:

„*1. Zvol jeden atribut jako kořen dílčího stromu,*

2. rozděl data v tomto uzlu na podmnožiny podle hodnot zvoleného atributu a přidej uzel pro každou podmnožinu,

3. existuje-li uzel, pro který nepatří data do téže třídy, pro tento uzel opakuj postup od bodu 1, jinak skonči.“

Volba vhodného atributu v prvním kroku může být určena dle charakteristik entropie, informačního zisku, poměrného informačního zisku, χ^2 nebo Giniho indexu (Berka, 2003, s. 87).

Algoritmus C4.5 představuje nástupce algoritmu ID3 se schopností zpracování číselných a chybějících hodnot, využívající jako kritérium dělení datové sady poměrný informační zisk (Rokach, 2005, s. 483):

$$pomerny_{info.zisk}(S,A) = \frac{zisk\ (S,A)}{pomerovy_{zisk}(S,A)} \tag{4.6}$$

$Zisk\ (S,A)$ představuje informační zisk, označující redukci entropie z nadřazeného uzlu (před rozdělením) na podřízené uzly (po rozdělení), přičemž $zisk\ (S,A)$ lze vyjádřit jako (Qui, 2017, s. 28):

$$zisk\ (S,A) = entropie(S) - \sum_{i=1}^{k} \frac{|S_i|}{|S|} entropie(S_i) \tag{4.7}$$

První část rovnice (4.7) vyjadřuje entropii původní sady S, druhá část představuje očekávanou hodnotu entropie po rozdělení S pomocí atributu A. Entropie pak v podstatě popisuje čistotu daných množin, k je počet rozdělení hodnot atributu a S_i je podmnožina instancí v i-tém potomku nadřazeného uzlu. $Pomerovy_{zisk}(S,A)$ představuje rozdělenou informaci vybraného dělícího atributu, který je definován jako (Lakshmi, 2013, s. 21; Qui, 2017, s. 28):

$$pomerovy_{zisk}(S,A) = -\sum_{i=1}^{k} \frac{|S_i|}{|S|} log_2 \frac{|S_i|}{|S|} \tag{4.8}$$

Proces výběru nového atributu a dělení datové sady se opakuje pro každý nekoncový uzel, přičemž atributy, které již byly zařazeny, jsou dále nevyužívány. V případě, kdy jsou všechny atributy vyčerpány anebo mají všechny trénovací vzory ve spojení s listem stejné cílové atributové hodnoty, je proces ukončen (Lakshmi, 2013, s. 22).

4.2.3. Podpůrné vektorové stroje

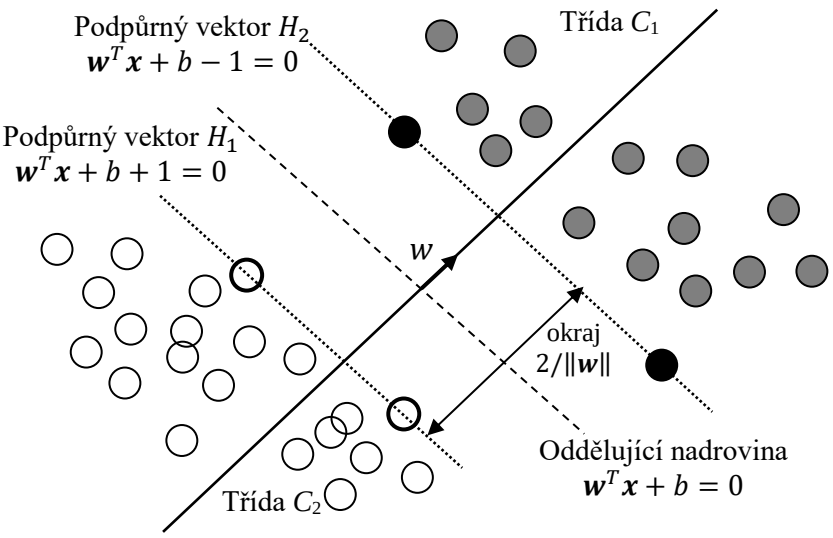
Metodika podpůrných vektorových strojů (SVM) představuje lineární diskriminační klasifikátor, s myšlenkou vytvoření nadroviny jako rozhodovací hranice mezi třídami během učení (Nicolas, 2014, s. 275). Oproti jiným klasifikátorům minimalizující empirické riziko, je zde rozdíl v podobě minimalizace rizika strukturálního (Pai, 2011, s. 315), přičemž oddělující nadrovina je zde vyjádřena rovnicí reprezentující model vygenerovaný prostřednictvím učení (Nicolas, 2014, s. 275).

Vzhledem k častým případům, kdy data nejsou v původním vstupním prostoru oddělitelná, umožňuje SVM také prostřednictví tzv. jádrové transformace (*kernel trick*) mapování dat do

prostoru vícerozměrného, kde jsou daná pozorování již oddělitelná snadněji (Russell, 2010, s. 744).

Také využití této metody se ukázalo jako vhodné pro detekci finančních podvodů s kreditními kartami a detekce spojené s podvody ve finančním výkaznictví. Učení je málo výpočetně náročné, avšak využití této metody může být limitující pro audity z důvodu obtížné interpretace výsledků v důsledku transformace vstupní sady (West, 2016, s. 52-61).

Za předpokladu lineárně oddělitelných dat mějme sadu trénovacích vzorů v podobě párů $(\boldsymbol{x}_i, y_i), i = 1, \dots, N$ kde $\boldsymbol{x_i} \in \mathbb{R}^d$ a $y_i \in \{\pm 1\}$. Mimo správné klasifikace je zde požadováno, aby se pozorování nacházely od nadroviny i v určité vzdálenosti (Alpaydin, 2010, s. 311), a tak SVM hledá optimální nadrovinu oddělující třídy tak, aby okraj definovaný jako vzdálenost mezi nejbližší instancí a nadrovinou byl maximální (Pai, 2011, s. 315-316). Bližší rozbor této problematiky je poukázán na níže uvedeném znázornění (Obrázek 13).



Obrázek 13: Podpůrné vektorové stroje – lineárně oddělitelné třídy

Zdroj: upraveno dle (Nicolas, 2014, s. 276)

Rovnice nadroviny, vyskytující ve středu hraničních bodů oddělující třídy $C_1(H_1: \boldsymbol{w}^T\boldsymbol{x} + b - 1)$ a $C_2(H_2: \boldsymbol{w}^T\boldsymbol{x} + b - 1)$ je definována lineární rovnicí $y = \boldsymbol{w}^T\boldsymbol{x} + b$. Roviny H_1 a H_2 pak představují podpůrné vektory. Okraj mezi podpůrnými vektory je zde pro všechna pozorování stejný, díky čemuž je tento případ označován jako „*hard margin*“. Optimalizační problém je pak definován jako (Nicolas, 2014, s. 276-277):

$$\min_{w,b} \left\{ \frac{\boldsymbol{w}^T\boldsymbol{w}}{2} \right\} \tag{4.9}$$

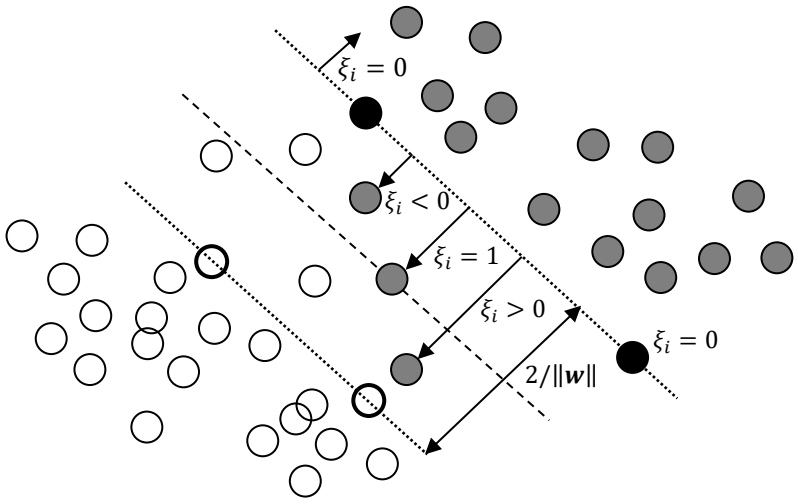
za podmínky $y_i(\boldsymbol{w}^T\boldsymbol{x} + b) \geq 1 \; \forall \; i$

V případě, kdy třídy nejsou lineárně oddělitelné, dochází k situaci, kdy podpůrné vektory v rámci učení nemohou daná pozorování zcela oddělit a stávají se tak lineárními funkcemi penalizující pozorování, která jsou umístěna chybně (Nicolas, 2014, s. 277). Definovány jsou zde tzv. doplňkové proměnné (*Slack Variables*) ξ, ukládající odchylku od okraje, kde případný výkyv může vzniknout v důsledku nesprávné klasifikace nebo nedostatečné vzdálenosti instance od okraje (Alpaydin, 2010, s. 315), zde označovaném jako „*soft margin*“. Grafické znázornění situace je uvedeno na následujícím vyjádření (Obrázek 14), přičemž optimalizační problém okraje je zde formulován (Nicolas, 2014, s. 277):

$$\min_{w,\xi}\left\{\frac{\boldsymbol{w}^T\boldsymbol{w}}{2} + C\sum_{i=0}^{n-1}\xi_i\right\} \tag{4.10}$$

za podmínek $\xi_i \geq 0, y_i(\boldsymbol{w}^T\boldsymbol{x} + b) \geq 1 - \xi \;\; \forall\, i$

kde C je faktor penalizace



Obrázek 14: Podpůrné vektorové stroje – lineárně neoddělitelné třídy

Zdroj: upraveno dle (Nicolas, 2014, s. 277)

Sekvenční minimální optimalizace (SMO) poté představuje iterační algoritmus pro učení SVM, jehož podstatu představuje problém kvadratického programování oddělující podpůrné vektory od ostatních trénovacích vzorů. Základem diskutovaného algoritmu je tak zjednodušení výše zmiňovaného kvadratického problému do jednodušších řešení, které je posléze možné analyticky řešit (Giusti, 2013, s. 57; Edwards, 2014, s. 37).

Prostřednictvím dvou Lagrangerových multiplikátorů poté algoritmus řeší optimalizační problém v každém kroku, kde po výběru těchto dvou multiplikátorů a výpočtu optimálních

hodnot dochází v SVM k aktualizaci, která nově vypočtenou hodnotu odráží (Edwards, 2014, s. 37).

Problém kvadratického programování je definován (Edwards, 2014, s. 37):

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} y_i y_j K(x_i x_j) \alpha_i \alpha_j \tag{4.11}$$

za podmínek

$$0 \leq \alpha_i \leq C, \forall\, i, \sum_{i=1}^{n} y_i \alpha_i = 0 \tag{4.12}$$

kde $y_i \in \{\pm 1\}$ a $\boldsymbol{x}_i$ je vstupní vektor, $K(x_i, x_j)$ představuje funkci jádra a C odkazuje na nadrovinu SVM. Proměnné α_i vyjadřují Lagrangeovy multiplikátory. Po rozpadu řešení na dílčí problémy je pro libovolné dva multiplikátory α_1 a α_2 omezení redukováno na (Edwards, 2014, s. 37):

$$0 \leq a_1, a_2 \leq C, y_1 a_1 + y_2 a_2 = k \tag{4.13}$$

Algoritmus SMO tak iterativně opakuje následující postup do té doby, dokud není dosaženo konvergence, přičemž při výběru proměnných α_i je použito heuristiky (Edwards, 2014, s. 37):

„*1. Nalezněte Lagrangerův multiplikátor α_1 porušující podmínky Karush–Kuhn–Tucker (KKT) pro optimalizační problém,*

2. zvolte druhý Lagrangerův multiplikátor α_2 a optimalizujte pár (α_1, α_2).“

4.3. Meta učící algoritmy

4.3.1. Bagging

Algoritmus bagging (*Bootstrap AGGregatING*) představuje metodiku spadající do oblasti meta učení s možností zlepšení přesnosti a stability klasifikačních a regresních modelů (Silva, 2010, s. 20). Diskutovaná problematika kombinuje rozhodnutí více modelů do jediné predikce, kde jsou výstupy z těchto modelů kombinovány prostřednictvím průměru (v případě regrese) anebo hlasováním (v případě klasifikace) (Witten, 2011, s. 352).

Podrobněji pak algoritmus vytváří různé vzory dat nahrazující původní trénovací sadu. Pro získání těchto různých podmnožin používá tato metoda vzorkování *bootstrap*, představující náhodné odebrání dat z původní datové množiny s možností opakovaného výskytu některých dat více než jednou a naopak (Marsland, 2015, s. 273). Opakováním výběrového procesu se tak získávají různé trénovací vzory pro každý klasifikační model, kde je výsledek konečného rozhodnutí klasifikace výsledkem hlasování každého člena klasifikátoru (Dua, 2011, s. 36),

neboli pro každý příklad je jeho konečná predikce třídy dána nejvyšším počtem předpovědí provedených základními modely. Prostřednictvím pseudokódu lze algoritmus popsat následovně (Aggarwal, 2014, s. 492-493):

Vstup: Trénovací sada $\mathcal{D} = \{\boldsymbol{x}_i, y_i\}_{i=1}^{m}$ $(\boldsymbol{x}_i \in \mathbb{R}^n, y_i \in Y)$

Výstup: Klasifikátor souboru H

- for $t \leftarrow 1$ to T do
- Vytvořte datovou množinu $\mathcal{D}_t$ náhodným odběrem vzorků nahrazující $\mathcal{D}$
- Naučte základní klasifikátor h_t založený na $\mathcal{D}_t$
- end for
- return $H(\boldsymbol{x}) = \arg\max_{y \in Y} \sum_{t=1}^{T} 1(h_t(\boldsymbol{x}) = y)$

Mimo zlepšení stability a přesnosti algoritmus redukuje rozptyl a napomáhá vyhnutí se přeučení, přičemž k využití tohoto algoritmu může být použit jakýkoliv typ modelu (Silva, 2010, s. 20).

4.3.2. Boosting

Boosting, stejně jako bagging, představuje metodiku kombinující stejné typy modelů, využívající ke kombinaci formu hlasování pro klasifikaci a průměr pro regresi. Rozdílnost oproti baggingu je zde vyjádřena v závislosti jednotlivých modelů, kdy je každý další model ovlivněn výkonem modelů předchozích (Witten, 2011, s. 358). Jednotlivé modely jsou zde tak konstruovány postupně, přičemž hlavní podstata metodiky je vyjádřena v generování několika relativně slabých klasifikátorů, mající pravděpodobnost chyby menší než 0.5 a jejich spojení do modelu silného, s pravděpodobností chyby libovolně malé (Silva, 2010, s. 19; Alpaydin, 2010, s. 431).

Nejvíce používanou variantou boostingu je algoritmus *AdaBoost* (Dua, 2011, s. 36). Trénovací vzory zde mají přidruženou váhu $w_j > 0$, přičemž čím vyšší je tato váha, tím vyšší je i důležitost vzoru. V počáteční fázi mají všechny trénovací vzory váhu stejnou $w_j = 1$, avšak po skončení klasifikace, kdy jsou některé vzory klasifikovány správně a některé ne, je u chybně klasifikovaných váha zvýšena a u správně klasifikovaných naopak snížena (Russell, 2010, s. 749).

Algoritmus pak při každé iteraci učí nové klasifikátory se sadou trénovacích vzorů, které jsou upraveny dle toho, jak úspěšně byly tyto vzory klasifikovány v minulosti (Marsland, 2015, s. 269), přičemž hlavní důraz je kladen na vzory, které byly klasifikovány chybně (Aggarwal,

2014, s. 496). Ve formě pseudokódu lze algoritmus vyjádřit následujícím způsobem (Marsland, 2015, s. 270):

Inicializace všech vah $^1/_N$, kde N je počet trénovacích vzorů

Dokud $0 < \varepsilon_t < \frac{1}{2}$(a $t < T$, T je maximální počet iterací):

- nauč klasifikátor $\{S, w^t\}$, získej hypotézu $h_t(x_n)$ pro data x_n
- vypočti chybu učení $\varepsilon_t = \sum_{n=1}^{N} w_n^{(t)} I(y_n \neq h_t(x_n))$
- vypočti váhy $\alpha_t = \log\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)$
- aktualizuj váhy pomocí: $w_n^{t+1} = w_n^t \exp(\alpha_t I(y_n \neq h_t(x_n))/Z_t$

  kde Z_t je normalizační konstanta

Výstup $f(x)$ =sign $(\sum_{t=1}^{T} \alpha_t h_t(x))$

4.3.3. Stacking

Algoritmus stacking (*Stacked Generalization*) představuje metodu, která na rozdíl od baggingu nebo boostingu, nepoužívá pro kombinaci stejné typy modelů, ale je postaven na učících se algoritmech, které jsou rozdílné. Jednotlivé modely jsou zde rozlišovány do dvou úrovní, a sice na modely tzv. nulté úrovně, které tvoří základní klasifikátory a model první úrovně, který je tvořen meta modelem propojující klasifikační modely z úrovně předešlé (Witten, 2011, s. 369).

Z důvodu zabránění přeučení v situaci, kdy by byla využita pro učení základních klasifikátorů ale i meta-klasifikátoru stejná trénovací sada, je vhodné data rozdělit na dvě podskupiny, přičemž nejčastějším přístupem je použití K násobné křížové validace (Aggarwal, 2014, s. 500). Modely v nulté úrovni jsou tak učeny a testovány, přičemž jejich výstupy jsou integrovány do souboru nové funkce, která je použita jako vstup do modelu vyšší úrovně. Po naučení modelu v první úrovni jsou pak všechny modely v nulté úrovni znovu naučeny pomocí celé trénovací sady (Kraipeerapun, 2015, s. 1290). Formou pseudokódu lze konečný algoritmus popsat následujícími kroky (Aggarwal, 2014, s. 501):

Vstup: Trénovací sada $\mathcal{D} = \{\boldsymbol{x}_i, y_i\}_{i=1}^{m}$ $(\boldsymbol{x}_i \in \mathbb{R}^n, y_i \in Y)$

Výstup: Klasifikátor souboru H

krok 1: Zavedení křížové validace trénovací sady pro první stupeň klasifikátoru

- Náhodně rozdělte $\mathcal{D}$ na podmnožiny K stejné velikosti $\mathcal{D} = \{\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_K\}$
- For $k \leftarrow 1$ to K do

krok 1.1: naučení klasifikátorů nulté úrovně

- for $t \leftarrow 1$ to T do
- Naučte klasifikátor h_{kt} z $\mathcal{D}\backslash\mathcal{D}_K$
- end for

krok 1.2: Vytvoření trénovací sady pro klasifikátor první úrovně

- for $x_i \in \mathcal{D}_K$ do
- Získejte vzor $\{\boldsymbol{x}_i^{\iota}, y_i\}$, kde $\boldsymbol{x}_i^{\iota} = \{h_{k1}(\boldsymbol{x}_i), h_{k2}(\boldsymbol{x}_i), \dots, h_{kT}(\boldsymbol{x}_i)\}$
- end for

- end for

krok 2: Naučení klasifikátoru první úrovně

- Naučte nový klasifikátor h^{ι} z podmnožiny $\{\boldsymbol{x}_i^{\iota}, y_i\}$

krok 3: Znovu naučte klasifikátor nulté úrovně

- for $t \leftarrow 1$ to T do
  - Naučte klasifikátor h_t na základě $\mathcal{D}$
- end for
- return $H(\boldsymbol{x}) = h^{\iota}\big(h_1(\boldsymbol{x}), h_2(\boldsymbol{x}), \dots, h_T(\boldsymbol{x})\big)$

5. DATA A NÁVRH MODELU

Následující část práce bude věnována popisu manipulaci s daty, která budou v nadcházející kapitole využita k návrhu klasifikačního modelu. Úvodní část kapitoly tak bude zaměřena na popis dat, jejich sběr a zhodnocení vybraných atributů z hlediska podvodů. Ve druhé části bude popsán použitý software, návrh modelu a sledované indikátory.

5.1. Data

Vzhledem k detekci podvodů v oblasti korporací, data pro tuto práci představují množinu dat společností, které se v rámci své činnosti podvodného jednání dopustily a společností, u kterých podvod detekován nebyl. Konečný datový soubor tak pochází z období let 2012-2015 a je tvořen veřejnými společnostmi působících na amerických burzách NYSE a NASDAQ. Podniky, které se podvodu dopustily, byly vybrány z veřejné databáze SEED[2], spolupracující s nezávislou vládní agenturou SEC, přičemž firmy dopouštějící se v uvedeném čtyřletém období klamání, tvoří v konečném počtu 80 společností pocházejících z různých průmyslových odvětví. Přehled těchto odvětví společně s přehledem o počtu podniků v nich působících je uveden v následujícím shrnutí (Tabulka 2).

Tabulka 2: Počet podniků vzhledem k průmyslovým odvětvím

průmyslové odvětví	počet podniků	průmyslové odvětví	počet podniků
bankovní a finanční služby	12	pojišťovnictví	2
těžební průmysl	9	informační služby	2
počítače/software/internet	8	investice a správa aktiv	2
farmaceutický průmysl	4	správa nemovitostí	2
zdravotnictví	4	maloobchod	2
inženýrství/výstavba	4	letectví a kosmonautika	1
elektronická zařízení	4	oděvní průmysl	1
obchodní a spotřebitelské služby	3	vzdělávání	1
zábavní průmysl	3	zemědělství	1
potravinářský průmysl	3	automobilový průmysl	1
telekomunikační zařízení	3	těžká konstrukce	1
výrobky pro domácnost	3	hotelnictví	1
chemický průmysl	2	elektrická zařízení	1

Zdroj: vlastní zpracování

U těchto vybraných společností byl pak kromě průmyslového odvětví zachycen také rok, kdy SEC nelegální činnost detekovala, přičemž nejčastější typ tvrzení, které Komise proti těmto společnostem vznesla, spadá do kategorií pochybení ve věci porušení ustanovení emitentního

[2] https://research.seed.law.nyu.edu/

vykazování informací a porušení zahraničního zákonu o korupčních praktikách. Jak je možné vidět z následujícího přehledu (Tabulka 3), obě výše uvedená tvrzení se pohybují ve všech čtyřech uvedených obdobích na prvních dvou pozicích ze všech jinak dále uvedených pochybení, přičemž tvrzení ve věci porušení emitentního vykazování informací nad ostatními pochybeními jasně dominuje.

Tabulka 3: Počet podniků vzhledem k typu pochybení (2012-2015)

rok	typ podvodu	počet podniků
2012	emitentní vykazování informací	7
	zahraniční zákon o korupčních praktikách	4
2013	emitentní vykazování informací	12
	zahraniční zákon o korupčních praktikách	6
	cenné papíry/veřejné penzijní fondy	1
	nabídka cenných papírů	1
2014	emitentní vykazování informací	18
	zahraniční zákon o korupčních praktikách	7
	makléř	1
	manipulace na trhu	1
2015	emitentní vykazování informací	11
	zahraniční zákon o korupčních praktikách	8
	jiné	2
	nabídka cenných papírů	1

Zdroj: vlastní zpracování

Protiklad k těmto institucím poté tvoří společnosti, které se podvodu nedopustily, přičemž je zde zachována podmínka výběru v podobě stejného průmyslového odvětví a podobné velikosti (tržní kapitalizace). Konečný počet této skupiny tvoří také 80 firem[3].

5.1.1. Sběr dat

Ke každé z výše sledovaných společností, byly následně dohledány příslušné finanční ukazatele, jejichž konkrétní výběr byl podřízen atributům používaných k detekci finančních podvodů na základě zjištění z několika vybraných studií. Konečný výčet těchto vybraných atributů je poukázán v níže uvedeném přehledu (Tabulka 4), pokrývajícím tak podstatné indikátory z oblasti účetní závěrky, neboť právě finanční výkazy představují vhodné finanční informace pro účelné finanční manipulace ze strany vedení podniku. Konkrétní hodnoty byly z podstatné části převzaty z webové databáze *MarketWatch*[3], přičemž vzhledem k roku, kdy SEC podvodné jednání u společností detekovala, byly informace k jednotlivým podnikům převzaty vždy za rok předchozí. V konečné podobě tak data představují 160 instancí (80 firem,

[3] https://www.marketwatch.com/

které se podvodu dopustily a 80, které nikoliv) a 35 vybraných atributů. Chybějící hodnoty byly označeny symbolem otazníku. Popisné statistiky k získaným atributům je možné vidět v příloze B, ukázku dat v příloze C.

Tabulka 4: Finanční atributy používané k detekci podvodů v oblasti finančních výkazů

finanční atribut	typ	reference
růst prodeje	číselný	(Abbasi, 2012, s. 1304)
růst pohledávek	číselný	(Abbasi, 2012, s. 1304)
růst čistého provozního toku z provozní činnosti	číselný	(Kotsiantis, 2006, s. 106)
závazky	číselný	(Li, 2014, s. 3552)
aktiva	číselný	(Li, 2014, s. 3552)
hrubý zisk	číselný	(Li, 2014, s. 3552)
čistý příjem	číselný	(Li, 2014, s. 3552)
celkové výnosy	číselný	(Li, 2014, s. 3552)
peněžní prostředky a peněžní ekvivalenty	číselný	(Li, 2014, s. 3552)
pohledávky	číselný	(Li, 2014, s. 3552)
peněžní tok z provozní činnosti	číselný	(Hoogs, 2007, s. 47)
peněžní tok z investiční činnosti	číselný	(Hoogs, 2007, s. 47)
provozní tok z finanční činnosti	číselný	(Hoogs, 2007, s. 47)
oběžná aktiva	číselný	(Hoogs, 2007, s. 47)
pracovní kapitál	číselný	(Hoogs, 2007, s. 47)
ukazatele rentability		
hrubý zisk/aktiva	číselný	(Li, 2014, s. 3552)
čistý zisk/aktiva	číselný	(Hoogs, 2007, s. 47)
čistý zisk/prodej	číselný	(Hoogs, 2007, s. 47)
čistý zisk/hrubý příjem	číselný	(Li, 2014, s. 3552)
hrubý zisk/prodej	číselný	(Li, 2014, s. 3552)
ukazatele likvidity		
pracovní kapitál/aktiva	číselný	(Hoogs, 2007, s. 47)
oběžná aktiva/krátkodobé závazky	číselný	(Li, 2014, s. 3552)
peněžní prostředky a peněžní ekvivalenty/aktiva	číselný	(Hoogs, 2007, s. 47)
zásoby/krátkodobé závazky	číselný	(Li, 2014, s. 3552)
hotovost/oběžná aktiva	číselný	(Li, 2014, s. 3552)
pohledávky/oběžná aktiva	číselný	(Hoogs, 2007, s. 47)
ukazatele aktivity		
zásoby/prodej	číselný	(Li, 2014, s. 3552)
prodej/aktiva	číselný	(Li, 2014, s. 3552)
struktura aktiv		
zásoby/aktiva	číselný	(Li, 2014, s. 3552)
oběžná aktiva/aktiva	číselný	(Li, 2014, s. 3552)
pohledávky/prodej	číselný	(Li, 2014, s. 3552)
pohledávky/aktiva	číselný	(Li, 2014, s. 3552)
dlouhodobý majetek/aktiva	číselný	(Ravisankar, 2011, s. 494)
ukazatele zadluženosti		
závazky/aktiva	číselný	(Li, 2014, s. 3552)
dlouhodobý dluh/aktiva	číselný	(Ravisankar, 2011, s. 494)

Zdroj: vlastní zpracování

5.1.2. Finanční atributy

S ohledem na problematiku finančních podvodů jsou podvody páchané v oblasti finančních výkazů, jak již bylo uvedeno v podkapitole 1.3.1, vyjádřeny v podobě nepravdivě odrážejícího stavu společnosti, za účelem zisku na úkor potenciálních investorů a bankovních institucí, poskytující těmto společnostem úvěr (Ravisankar, 2011, s. 491).

V oblasti *výkazu zisku a ztráty* je tedy zřejmé, že k záměrné manipulaci dochází nejčastěji u finančních údajů týkajících se výnosů a nákladů, neboť se jedná o způsob, jak nejlépe přímo a v krátké době ovlivnit finanční výsledky podniku (Dimitrijevic, 2015b, s. 136). Společnosti dopouštějící se podvodů v této oblasti tak často přidávají fiktivní výnosy, čímž zapříčiní i jejich zvýšení (Abbasi, 2012, s. 1304) a právě abnormální hodnota tohoto ukazatele může poukázat na výskyt případného podvodu (Ravisankar, 2011, s. 493). V otázce *rozvahy* lze zmínit častý způsob klamání v podobě nadhodnocení pohledávek, manipulaci se zásobami, nadhodnocení a podhodnocení závazků nebo jejich záměrné opomenutí. Atributy z oblasti *přehledu o peněžních tocích* pak bývají méně častým předmětem manipulace, avšak metody manipulující s operačními aktivitami lze obecně vyjádřit v podobě metod maximalizující příliv peněz z provozní činnosti, kde mohou být vedením do oblasti záměrně uvedeny položky patřící do oblasti finančních aktivit anebo metod minimalizující odliv peněz (Dimitrijevic, 2015b, s. 138-143).

Podstatnou část vybraných atributů poté představují finanční poměrové ukazatele, které jsou rozděleny do pěti kategorií, a sice na ukazatele *rentability*, *likvidity*, *aktivity*, *struktury aktiv* a ukazatele *zadluženosti*. Poměrové ukazatele rentability poukazují na schopnost společnosti generovat návratnost svých zdrojů (Ravisankar, 2011, s. 492), přičemž nižší ziskovost může vést k nadhodnocení výnosů nebo podhodnocení výdajů (Dalnial, 2014, s. 64). Obdobný případ naznačují i ukazatele likvidity zaměřující se na schopnost podniků splácet krátkodobé závazky (Ravisankar, 2011, s. 492), kdy čím nižší je likvidita, tím spíše bude vedení k podvodu motivováno (Dalnial, 2014, s. 64). Opačný přístup je veden v oblasti ukazatelů zadluženosti, kdy je zřejmé, že vyšší pákový efekt představuje větší riziko k porušení úvěrových smluv, což může vést k podhodnocení závazků a aktiv (Persons, 2011, s. 40). Ukazatelé aktivity vypovídají o správě majetku společnosti (Ravisankar, 2011, s. 492), přičemž v případě struktury aktiv lze poukázat především na položky posuzující podvody v podobě pohledávek a zásob, jejichž nadhodnocení může být zřetelným podvodným ukazatelem (Dalnial, 2014, s. 64).

5.2. Použitý software

Pro účely experimentů bude použit nástroj Weka 3.8.2. (*Waikato Environment for Knowledge Analysis*), obsahující soubor algoritmů strojového učení a nástrojů pro předzpracování dat. Program vznikl na univerzitě Waikato na Novém Zélandu a pracuje téměř na libovolné platformě. Je napsán v jazyce Java a distribuován po licencí GNU General Public Licence (Witten, 2011, s. 403-404).

5.3. Návrh modelu

Z důvodů relativně většího počtu atributů vzhledem k počtu datových vzorků byla v dalším kroku provedena selekce těch atributů, které budou pro učení nejdůležitější. Tento způsob tak sníží dimenzionalitu dat, umožní rychlejší a efektivnější učení a může zajistit i vyšší přesnost klasifikace (Hall, 1999, s. 25). V oblasti přípravy dat tak byla množina dat pětkrát náhodně zamíchána, přičemž na každý takto vzniklý vzorek bylo aplikováno rozdělení na trénovací a testovací část v poměru 4:1 (80 % trénovací & 20 % testovací). Výběr atributů byl proveden prostřednictví metody hodnocení *CfsSubsetEval* a vyhledávací metody *BestFirst*.

Hodnotitel *CfsSubsetEval* je založen na korelaci, přičemž vyhodnotí podmnožinu atributů tím, že zváží individuální prediktivní schopnost jednotlivých atributů spolu se stupněm redundance mezi nimi. Vybrané atributy jsou poté takové, které jsou vysoce korelovány s výstupní třídou a zároveň nejsou korelovány s atributy jinými (Wallace, 2010, s. 135). Formálně lze tento hodnotitel vyjádřit následovně (Hall, 1997, s. 2):

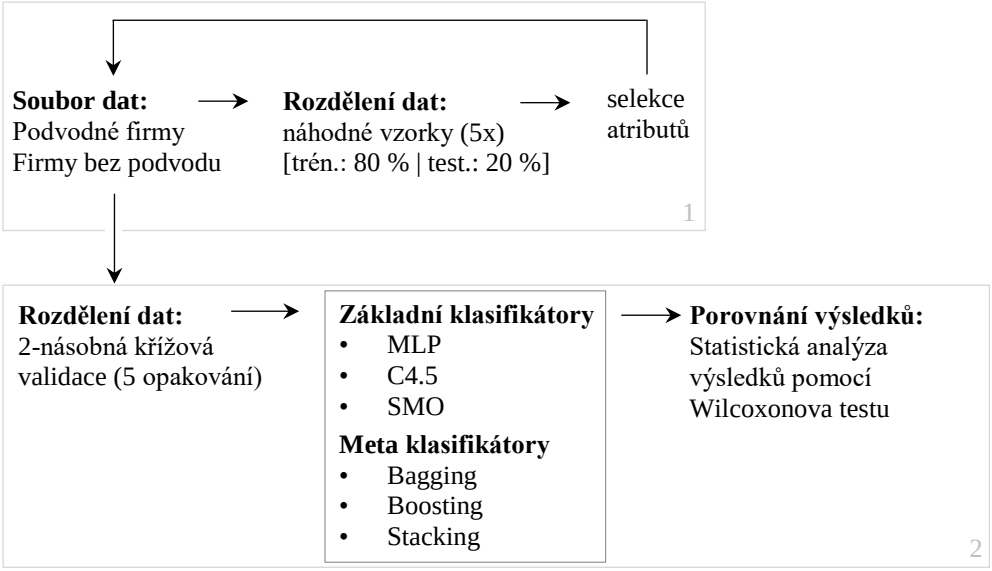
$$G_s = \frac{k\bar{r}_{ci}}{\sqrt{k + k(k-1)\overline{r_{ii'}}}} \tag{5.1}$$

kde k je počet atributů v podmnožině, $\bar{r}_{ci}$ je průměrná korelace atributu se třídou a $\bar{r}_{ii'}$ je průměrná korelace mezi atributy.

Metoda *BestFirst* poté prohledává prostor podmnožin atributů pomocí hladového algoritmu s možností zpětného vyhledávání (Wallace, 2010, s. 135). Algoritmus tak vytváří nové podmnožiny na základě přidání nebo odebrání atributů do podmnožiny stávající, avšak je zde možnost návratu podél cesty výběru v hledání jiných možností, v případě výskytu aktuálního, již nezlepšujícího řešení (Williams, 2006, s. 9).

Proces selekce byl aplikován na trénovacích datech, přičemž na každém z pěti vzorků bylo vybráno sedm nejdůležitějších atributů: *růst prodeje*, *růst pohledávek*, *hrubý zisk*, *čistý příjem*, *peněžní prostředky a peněžní ekvivalenty*, *oběžná aktiva* a *poměr čistého zisku k prodeji*. Na takto zredukovaný výběr atributů byla následně použita 2-násobná křížová validace s pěti

opakováními a poté aplikovány zvolené metody v podobě základních klasifikátorů (neuronové sítě (MLP), rozhodovacího stromu (C4.5), podpůrných vektorových strojů (SMO)) a meta klasifikátorů (*Bagging*, *Boosting* a *Stacking*). Grafický návrh celkového modelu je poukázán na následujícím zobrazení (Obrázek 15).



Obrázek 15: Návrh modelu klasifikace finančních podvodů

Zdroj: vlastní zpracování

5.4. Sledované ukazatele

Ke zhodnocení výsledků experimentů budou následně sledovány čtyři ukazatele, a sice *procentuální úspěšnost* správně klasifikovaných případů, *senzitivita*, *specifita* a *plocha pod ROC křivkou*. Zvolené indikátory vycházejí z matice záměn (Tabulka 5), jejíž vyjádření je definováno ve čtvercové podobě obsahující všechny použité třídy (Marsland, 2015, s. 21).

Tabulka 5: Matice záměn pro dvě třídy

	předpokládaná třída		
aktuální třída	pozitivní	negativní	celkem
pozitivní	*tp*: skutečně pozitivní	*fn*: falešně negativní	*p*
negativní	*fp*: falešně pozitivní	*tn*: skutečně negativní	*n*
celkem	*p´*	*n´*	*N*

Zdroj: upraveno dle (Alpaydin, 2010, s. 489)

Skutečně pozitivní (*tp*) a skutečně negativní (*tn*) klasifikace, vyskytující se na hlavní diagonále matice, vyjadřuje správnou odpověď systému. Falešně pozitivní (*fp*) je takové vyjádření, kdy je třída nesprávně předpovězena jako pozitivní, ale ve skutečnosti je negativní. Falešně negativní (*fn*) pak naopak představuje výsledek, který je nesprávně předpovězen jako negativní, ale ve skutečnosti je pozitivní (Witten, 2011, s. 164; Marsland, 2015, s. 21-22).

Celková míra úspěšnosti neboli přesnost (*accuracy*) klasifikace vyjadřuje procento správně klasifikovaných příkladů z celkového počtu příkladů (West, 2015, s. 464):

$$presnost = (tp + tn)/N \tag{5.2}$$

Senzitivita (*sensitivity*), jinak také skutečná pozitivní míra (*tp_rate*) nebo úplnost (*recall*), vyjadřuje procento správně klasifikovaných pozitivních vzorků z celkového počtu všech pozitivních vzorků. V tomto případě senzitivita zastupuje firmy, které se podvodu nedopustily (West, 2015, s. 464):

$$senzitivita = tp/p \tag{5.3}$$

Specificita neboli skutečná negativní míra (*tn_rate*) pak naopak vyjadřuje procento počtu správně klasifikovaných negativních případů z celkového počtu všech negativních vzorků. V této práci zastupuje podniky, u kterých podvod detekován byl (West, 2015, s. 464):

$$specifita = tn/n \tag{5.4}$$

ROC křivka vyjadřuje dvourozměrné grafické znázornění výkonu klasifikátorů (Fawcett, 2006, s. 868), ovšem v případě, kdy je potřebné redukovat ROC křivku do jediné hodnoty, lze využít plochu pod křivkou (AUC), nabývající v ideálním případě hodnotu jedné (Alpaydin, 2010, s. 491). Hodnota AUC má poté důležitou statistickou vlastnost, a sice že odpovídá pravděpodobnosti, že klasifikátor vyhodnotí náhodně vybraný pozitivní vzor lépe než náhodně zvolený negativní vzor (Fawcett, 2006, s. 868).

6. PROVEDENÍ EXPERIMENTŮ

Obsah této části práce je zaměřen na měření a zhodnocení získaných výsledků, přičemž veškeré výsledky jsou zde uvedeny v podobě průměru z celkem 10 běhů experimentů (podrobnější výsledky jsou uvedeny v příloze E). V rámci měření bylo dále vyzkoušeno několik kombinací nastavení parametrů u jednotlivých algoritmů strojového učení. Přehled těchto nastavení je uveden v příloze D.

V následujícím přehledu (Tabulka 6) jsou poté uvedena nastavení, která se v rámci přesnosti klasifikace jevila jako nejlepší tzn. že při těchto nastavení bylo dosaženo nejvyšší přesnosti správně klasifikovaných případů a odpovídají tak výsledkům experimentů uvedených dále.

Tabulka 6: Nastavení parametrů metod strojového učení

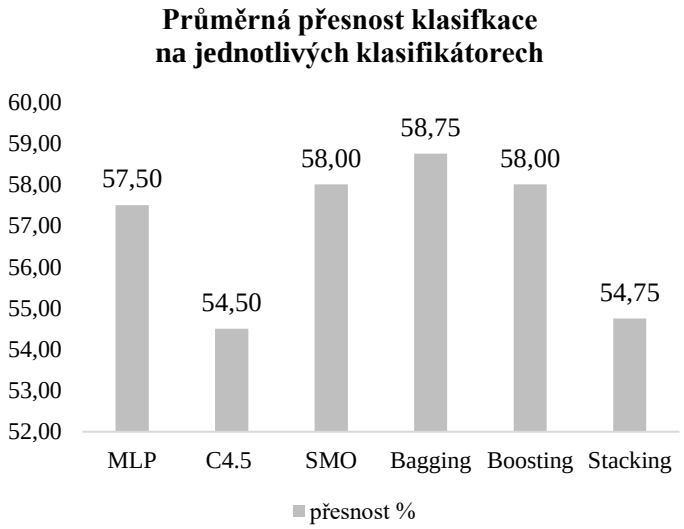
klasifikátor	nastavení parametrů
MLP	trénovací čas: 500, neurony ve skryté vrstvě: 9, momentum: 0.4, rychlost učení: 0.2
C4.5	faktor spolehlivosti: 0.25, minimální počet vzorů na list: 20
SMO	jádrová funkce: polyKernel, exponent: 1, parametr komplexnosti: 31
Bagging	klasifikátor: C4.5, počet iterací: 100, velikost bagu jako procento z trénovacích dat: 70
Boosting	klasifikátor: SMO, jádrová funkce: RBFkernel, gamma: 0.2, parametr komplexnosti: 54
Stacking	klasifikátor = {MLP, C4.5}, meta klasifikátor: SMO = {jádrová funkce: polyKernel, exponent: 1, parametr komplexnosti: 15}

Zdroj: vlastní zpracování

6.1. Výsledky experimentů

Následující znázornění (Obrázek 16) vyjadřuje porovnání výsledků experimentů z pohledu měřítka průměrné míry úspěšnosti na jednotlivých klasifikátorech. Z grafického vyhodnocení je viditelné, že nejlepšího výsledku bylo dosaženo při použití meta klasifikátoru *Bagging*, jehož úspěšnost jako jediná přesáhla hranici 58,00 %.

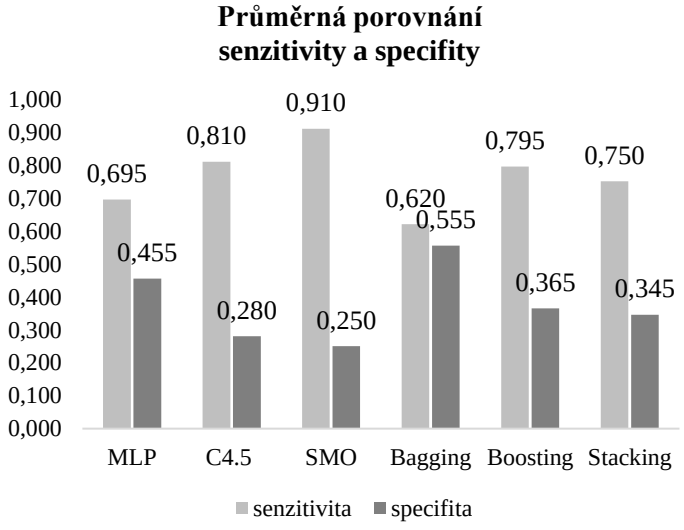
V případě metod SMO a *Boosting* bylo dosaženo průměrné přesnosti klasifikovaných případů se shodnou úspěšností, a sice 58,00 %. Naopak nejhůře v průběhu experimentů dopadla metoda rozhodovacího stromu C4.5, kde průměrná přesnost správně klasifikovaných případů dosáhla úspěšnosti 54,50 %. Z pohledu porovnání základních klasifikátorů a meta klasifikátorů lze říci, že nejlépe v oblasti základních klasifikátorů dopadl algoritmus SMO a v případě meta klasifikátorů *Bagging*, přičemž právě meta klasifikátory dosáhly v porovnání se základními klasifikátory v celkovém v průměru lepší přesnosti o 0,5 %.



Obrázek 16: Průměrná přesnost klasifikace

Zdroj: vlastní zpracování

V následujícím grafu (Obrázek 17) jsou ukázána srovnání výsledků z pohledu měřítka počtu společností, které se podvodu dopustily a byly správně předpovězeny jako podvodné (*specificita*) a společností, které byly správně klasifikovány jako nepodvodné z celkového počtu všech skutečně nepodvodných společností (*senzitivita*). V případě specificity je zřejmé, že nejlépe byl vyhodnocen algoritmus *Bagging*, neboť specificita zde dosahuje průměrné hodnoty až 0,555. Na druhém místě se umístil algoritmus MLP (0,455) a naopak nejhůře zde dopadl algoritmus SMO, jehož průměrná hodnota správně klasifikovaných podvodných společností dosáhla hodnoty 0,250. V případě senzitivity jsou výsledky, mimo algoritmů *Boosting* a *Stacking*, kde je pozice umístění prohozena, opačné.

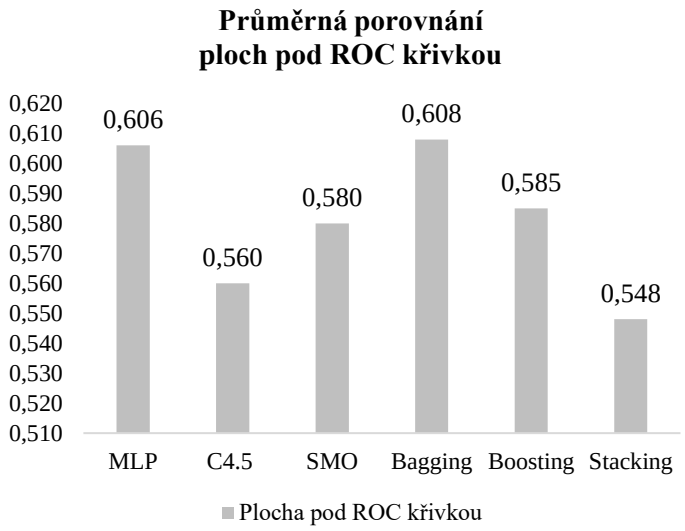


Obrázek 17: Průměrná porovnání senzitivity a specificity

Zdroj: vlastní zpracování

Dále si je možné povšimnout, že senzitivita ve všech případech měření převyšuje specificitu, tzn. že algoritmy měly problém klasifikovat ty firmy, které se podvodu dopustily. Toto zjištění je tedy spíše nežádoucí, a ačkoliv je senzitivita vyšší než specificita u všech algoritmů, je jasně zřetelné, že algoritmus *Bagging* byl co se týká poměru senzitivity a specificity jako jediný vyrovnaný.

Poslední grafické vyjádření (Obrázek 18) znázorňuje účinnost klasifikace z pohledu plochy pod ROC křivkou (AUC). Je možné si povšimnout, že i zde dosáhla měření nejlepší průměrné účinnosti meta klasifikátor *Bagging* (0,608), druhé v pořadí byly vyhodnoceny neuronové sítě MLP (0,606). Algoritmus SMO a *Boosting* se pohybují opět téměř ve shodné účinnosti, přičemž nejhůře zde byl naopak vyhodnocen meta klasifikátor *Stacking*, jehož úspěšnost klasifikace dosáhla průměrné hodnoty 0,548.



Obrázek 18: Průměrná hodnocení plochy pod ROC křivkou

Zdroj: vlastní zpracování

Z pohledu zhodnocení všech ukazatelů a použitých algoritmů jsou výsledky měření poměrně shodné, avšak z celkových šesti použitých metod lze vyhodnotit jako nejlepší metodu meta klasifikátor *Bagging*, jenž na ukazateli celkové míry úspěšnosti dosáhl nejlepšího výsledku ze všech jinak dále použitých algoritmů, a to 58,75 %. Nejlepší výsledek byl poté zaznamenán i v případě specificity (0,555) a plochy pod ROC křivkou (0,608).

Na druhém místě se z pohledu přesnosti klasifikace umístil shodně algoritmus SMO a meta klasifikátor *Boosting*, kteří dosáhli shodné přesnosti (58,00 %), a těsně za nimi pak algoritmus MLP (57,50 %). Z pohledu plochy pod ROC křivkou lze ovšem spíše uvažovat jako druhý v pořadí algoritmus MLP a poté *Boosting*.

6.2. Statistické zhodnocení výsledků

Z pohledu statistického zhodnocení výsledků byl použit neparametrický Wilcoxonův párový test (nevyžaduje předpoklad normálního rozdělení), jehož vyhodnocení bylo provedené prostřednictvím programu Statistica 12.

Pro zhodnocení naměřených hodnot v porovnání s jednotlivými sledovanými indikátory a použitými klasifikačními algoritmy tak byla stanovena hypotéza H_0, kdy *se předpokládá, že výkonnost jednotlivých metod vzhledem ke stanoveným měřítkám není statisticky významně rozdílná* a alternativní hypotéza H_1, kdy *se předpokládá, že výkonnost jednotlivých metod vzhledem ke stanoveným měřítkám je statisticky významně rozdílná*. V následujících čtyřech přehledech jsou poté uvedeny výsledky z pohledu jednotlivých měřítek a použitých metod, při různých hladinách významnosti.

V prvním případě (Tabulka 7) je možné vidět, že statisticky významný rozdíl byl zaznamenán u algoritmu *Stacking*, jenž byl vyhodnocen jako statisticky významně horší než metoda SMO, *Boosting* a *Bagging*. Další statisticky významný rozdíl byl zaznamenán i mezi metodami C4.5 a MLP, kde právě metoda rozhodovacího stromu vzešla z uvedeného porovnání statisticky hůře.

Tabulka 7: Test významnosti klasifikátorů – přesnost

	MLP	C4.5	SMO	Bagging	Boosting
C4.5	0,093#				
SMO	0,610	0,193			
Bagging	0,441	0,155	0,726		
Boosting	1,000	0,221	0,813	0,683	
Stacking	0,286	0,760	0,086#	0,037##	0,067#

Wilcox. párový test: * statisticky lepší na hladině významnosti $p<0.1$, ** statisticky lepší na hladině významnosti $p<0.05$, # statisticky horší na hladině významnosti $p<0.1$, ## statisticky horší na hladině významnosti $p<0.05$

Zdroj: vlastní zpracování

U ukazatele senzitivity (Tabulka 8) jsou výsledky při nižší hladině významnosti zřetelnější, avšak z pohledu celkového shrnutí je zde jasně viditelná převaha algoritmu SMO nad všemi ostatními metodami. Naopak *Bagging* zde byl vyhodnocen jako významně horší, a to nejen proti metodě SMO, ale také proti algoritmům C4.5 a *Boosting*.

Tabulka 8: Test významnosti klasifikátorů– senzitivita

	MLP	C4.5	SMO	Bagging	Boosting
C4.5	0,114				
SMO	0,009**	0,086*			
Bagging	0,386	0,008##	0,005##		
Boosting	0,262	0,646	0,050##	0,014**	
Stacking	0,260	0,508	0,093#	0,114	0,445

Wilcox. párový test: * statisticky lepší na hladině významnosti $p<0.1$, ** statisticky lepší na hladině významnosti $p<0.05$, # statisticky horší na hladině významnosti $p<0.1$, ## statisticky horší na hladině významnosti $p<0.05$

Zdroj: vlastní zpracování

V případě specificity (Tabulka 9), stejně jako u senzitivity, jsou i zde mezi metodami zaznamenány při nižší hladině významnosti statisticky významné rozdíly, ovšem opět z pohledu celkového shrnutí je zde naopak zřetelně viditelná významná převaha algoritmu *Bagging*, jenž byl vyhodnocen, kromě metody MLP, jako statisticky významně lepší než ostatní algoritmy. Dalšího statisticky lepšího výsledek zde zaznamenala i neuronová síť, jejíž převaha byla zaznamenána proti metodám C4.5, SMO a metodě *Stacking*.

Tabulka 9: Test významnosti klasifikátorů – specificita

	MLP	C4.5	SMO	Bagging	Boosting
C4.5	0,028##				
SMO	0,013##	0,575			
Bagging	0,203	0,017**	0,008**		
Boosting	0,445	0,386	0,097*	0,015##	
Stacking	0,093#	0,508	0,343	0,093#	0,445

Wilcox. párový test: * statisticky lepší na hladině významnosti $p<0.1$, ** statisticky lepší na hladině významnosti $p<0.05$, # statisticky horší na hladině významnosti $p<0.1$, ## statisticky horší na hladině významnosti $p<0.05$

Zdroj: vlastní zpracování

Poslední statistické zhodnocení je zde uvedeno z pohledu ukazatele pod ROC křivkou (Tabulka 10). Na tomto místě jsou výsledky Wilcoxonova testu při nižší hladině významnosti opět méně zřetelnější, avšak i zde je významně horší výsledek zaznamenán u metody *Stacking*, jenž byl významně překonán téměř všemi algoritmy s výjimkou metody C4.5.

Tabulka 10: Test významnosti klasifikátorů – plocha pod ROC křivkou

	MLP	C4.5	SMO	Bagging	Boosting
C4.5	0,169				
SMO	0,575	0,441			
Bagging	0,878	0,114	0,203		
Boosting	0,386	0,308	0,386	0,285	
Stacking	0,051#	0,838	0,086#	0,007##	0,028##

Wilcox. párový test: * statisticky lepší na hladině významnosti $p<0.1$, ** statisticky lepší na hladině významnosti $p<0.05$, # statisticky horší na hladině významnosti $p<0.1$, ## statisticky horší na hladině významnosti $p<0.05$

Zdroj: vlastní zpracování

Z konečného shrnutí si lze tak povšimnout statisticky významné převahy algoritmu *Bagging*. Z pohledu specifity lze ovšem zmínit i významně lepší výsledek ze strany neuronové sítě a v případě senzitivity pak metodu SMO. Statisticky horších výsledků zde naopak zaznamenal algoritmus *Stacking*.

Z výše uvedeného zle tak učinit konečný závěr, kdy se zamítá hypotéza H_0 a přijímá se hypotéza H_1, a tedy že *výkonnost jednotlivých metod vzhledem ke stanoveným měřítkám* ***je statisticky významně rozdílná***.

ZÁVĚR

Diplomová práce představuje pojednání o detekci finančních podvodů metodami strojového učení, přičemž cíl práce představoval dílčí kroky v podobě zabezpečení sběru dat, jejich popsání, charakteristice vybraných algoritmů strojového učení, navrhnutí modelu pro detekci finančních podvodů a provedení porovnání výsledků zvolených metod.

Struktura práce je tvořena čtyřmi základními oblastmi, a sice kdy je v první části shrnuta problematika finančních podvodů z hlediska pojmů, dělení, preventivních opatření a detekčních technik. Druhá část je zaměřena na shrnutí výsledků ze zprávy Komise pro cenné papíry a burzy (SEC) z období let 2010-2015 a třetí část na teoretické základy strojového učení a vybrané algoritmy. Klíčovou podstatu v rámci plnění cíle ovšem představuje část čtvrtá, týkající se dat a návrhu modelu pro detekci finančních podvodů. Data pro tuto práci tak představovala finanční údaje získané z finančních výkazů z období let 2012-2015 společností působících na amerických burzách NYSE a NASDQ, které se dopustily nebo nedopustily podvodu.

V rámci splnění cíle bylo nutné potřebná data sesbírat, přičemž výsledná množina dat tak byla z podstatné části převzata z webové databáze *MarketWatch*, kde byl výběr konkrétních finančních indikátorů podřízen výběru atributů používaných k detekci na základě zjištění z několika studií zabývající se touto problematikou. Z důvodu získaní rozsáhlejšího výběru finančních atributů vzhledem k počtu podniků, byl v dalším kroku proveden výběr těch atributů, které budou znamenat pro výsledky klasifikace větší predikční schopnost. Na množinu dat tak byl aplikován filtr výběru, jenž vyhodnotil na základě průměrné predikční schopnosti sedm důležitých atributů z celkových třiceti pěti. S takto zredukovanou datovou množinou bylo v dalším kroku provedeno měření se zvolenými metodami v podobě neuronové sítě (MLP), rozhodovacího stromu (C4.5), podpůrných vektorových strojů (SMO) a meta učících metod: *Bagging*, *Boosting* a *Stacking*. Výsledky experimentů byly posouzeny z pohledu čtyř měřítek, a sice z pohledu přesnosti klasifikace, senzitivity, specificity a plochy pod ROC křivkou.

V konečném výsledku byl poté jako nejlepší metoda pro detekci finančních podvodů vyhodnocen meta klasifikátor *Bagging*, jenž dosáhl na měřítkách přesnosti, specificity a plochy pod ROC křivkou nejlepších výsledků. V posledním kroku bylo následně provedeno testování získaných výsledků při použití Wilcoxonova párového testu, jenž vyhodnotil rozdíly ve sledovaných měřítkách vzhledem k použitým metodám jako statisticky významně rozdílné.

Z pohledu zhodnocení získaných měření lze ovšem konstatovat jistý rozdíl oproti jiným studiím, kde byl v několika případech zaznamenán lepší výsledek na přesnosti klasifikovaných případů. Tuto skutečnost si lze vysvětlit možným vyváženým poměrem počtu vzorků

sledovaných společností, neboť mnozí autoři (Kotsiantis, 2006, s. 106-109; Pai, 2011, s. 317-319; Li, 2014, s. 3551-3554) používají nízké zastoupení podvodných firem, oproti těm, které se podvodu nedopustily. Přesnost klasifikace pak u těchto autorů činí v případě algoritmu MLP (75,44 % - 82,67 %), C4.5 (82,46 % - 91,20 %) a SVM (78,66 % - 92,00 %). V porovnání s jinými autory (Ravisankar, 2011, s. 494-497; Chen, 2014a, s. 4-6; Chen, 2014b, s. 225-232), kteří mají zastoupení v jednotlivých třídách vyvážené, lze zaznamenat v přesnosti klasifikace jisté rozdíly: MLP (67,09 % - 78, 77 %), C4.5 (79,00 % - 85,70 %), SVM (72,02 % - 73,41 %). Bylo by proto zajímavé věnovat vlivu poměru zastoupení podniků v jednotlivých třídách v budoucnosti pozornost.

Další jistou nevýhodou může být i skutečnost, že detekce z finančních výkazů proběhla pro několik různých typů podvodů, což se může jevit v případě snahy o realizaci detekčního modelu jako věc obtížná. Spolehlivější a patrně i lepší výsledky by byly zaznamenány, kdyby se model detekce soustředil pouze na jeden konkrétní typ pochybení.

Jak již bylo uvedeno při popisu jednotlivých algoritmů, základní klasifikátory byly vybrány v souladu s používanými technikami zabývající se touto problematikou (West, 2016, s. 60; Ngai, 2011, s. 564), v případě použití meta klasifikátorů pak bylo očekáváno lepších výsledků klasifikace (Kotsiantis, 2006, s. 109; Song, 2014, s. 621). Tento předpoklad byl nakonec splněn, neboť meta učící algoritmy dosahovali na měřítku přesnosti klasifikace v průměru lepšího výsledku než algoritmy základní, a proto může být v budoucnu přístup užití meta klasifikátorů společně s nevyváženým počtem podniků v jednotlivých třídách slibným řešením.

Závěrem lze zmínit, že cíl diplomové práce byl splněn a tato práce tak může být vhodným základem pro další rozvoj k řešení problému odhalování finančních podvodů. Budoucím rozšířením by mohlo být zaměření se na větší časový interval z důvodu potvrzení získaných výsledků.

POUŽITÁ LITERATURA

1. ABBASI, Ahmed, 2012. MetaFraud: A Meta-Learning Framework for Detecting Financial Fraud. *MIS quarterly*. **36**(4), 1293-1327. ISSN 02767783.
2. ACFE, 2016. *Report to the nations on occupational fraud and abuse: 2016 Global fraud study* [online]. 92 s. [cit. 2017-10-20]. Dostupné z: https://s3-us-west-2.amazonaws.com/acfepublic/2016-report-to-the-nations.pdf
3. AGGARWAL, Charu C., 2014. *Data classification: algorithms and applications*. Boca Raton: CRC Press, Taylor & Francis Group. ISBN 9781466586741.
4. ALBRECHT, W. Steve, Chad O. ALBRECHT, Conan C. ALBRECHT a Mark F. ZIMBELMAN, 2011. *Fraud Examination*. 4th ed. Mason OH: South Western, Cengage Learning. ISBN 0-538-47084-4.
5. ALFORD, Mike, 2013. Intelligent fraud detection: a comparison of neural and Bayesian methods. *Computer fraud & security*. **2013**(4), 14-16. ISSN 13613723.
6. ALPAYDIN, Ethem., 2010. *Introduction to machine learning*. 2nd ed. Cambridge, Mass.: MIT Press. Adaptive computation and machine learning. ISBN 978-0-262-01243-0.
7. BEASLEY, Mark S., Joseph V. CARCELLO a Dana R. HERMANSON, 1999. *Fraudulent Financial Reporting: 1987-1997 An Analysis of U.S. Public Companies: Committee of Sponsoring Organizations of the Treadway Commission*.
8. BERKA, Petr, 2003. *Dobývání znalostí z databází*. Vyd. 1. Praha: Academia. ISBN 80-200-1062-9.
9. BUNT, Harry C., John CARROLL a Giorgio. SATTA, 2004. *New developments in parsing technology*. Boston: Kluwer Academic Publishers. ISBN 140202293x.
10. CARMICHAEL, D. R., Ray WHITTINGTON a Lynford. GRAHAM, 2007. *Accountants' handbook*. 11th ed. Hoboken, N.J.: Wiley. ISBN 9780471790396.
11. CECCHINI, Mark, 2010. Detecting Management Fraud in Public Companies. *Management science*. **56**(7), 1146-1160. ISSN 00251909.
12. COENEN, Tracy, 2008. *Essentials of corporate fraud*. 1st. Hoboken, N.J.: John Wiley & Sons. Essentials series. ISBN 978-0-470-19412-6.
13. COSO, 2013. *Internal Control-Integrated Framework: Executive Summary* [online]. [cit. 2017-11-03]. ISBN 978-1-93735-239-4. Dostupné z: www.coso.org/Documents/990025P-Executive-Summary-final-may20.pdf
14. CRUZ, Marcelo G., Gareth V. PETERS a Pavel V. SHEVCHENKO, 2015. *Fundamental Aspects of Operational Risk and Insurance Analytics: A Handbook of Operational Risk*. 1st. Hoboken, New Jersey: Wiley. Wiley handbooks in financial engineering and econometrics. ISBN 978-1-118-11839-9.
15. DALNIAL, Hawariah, Amrizah KAMALUDDIN, Zuraidah Mohd SANUSI a Khairun Syafiza KHAIRUDDIN, 2014. Accountability in Financial Reporting: Detecting Fraudulent Firms. *Procedia, social and behavioral sciences*. **145**, 61-69. ISSN 18770428.
16. DIMITRIJEVIC, Dragomir, 2015b. The detection and prevention of manipulations in the balance sheet and the cash flow statement. *Ekonomski horizonti*. **17**(2), 137-150. ISSN 1450863X.
17. DIMITRIJEVIC, Dragomir, Vesna MILOVENCOVIC a Vladimir STANCIC, 2015a. The Role of A Company's Internal Control System in Fraud Prevention. *E-finanse*. **11**(3), 34-44. ISSN 1734039X.

18. DORMINEY, Jack, Scott A. FLEMING, Mary-Jo KRANACHER a Richard A. RILEY, 2012. The Evolution of Fraud Theory. *Issues in accounting education*. **27**(2), 555-579. ISSN 07393172.
19. DUA, Sumeet a Xian DU, 2011. *Data Mining and Machine Learning in Cybersecurity*. Hoboken: CRC Press. ISBN 9781439839430.
20. DUBIS, Gregory S., Abraham D. AKRESH, Princy JAIN, Lynn MORLEY, Theresa M. PHIPPS a Richard A. SCHMIDT, 2009. *Internal Auditing and Fraud: IPPF – Practice Guide* [online]. 42 s. [cit. 2017-10-20]. Dostupné z: www.kcgaudit.co.uk/wp-content/uploads/2016/01/internal_auditing_and_fraud.pdf
21. DVOŘÁČEK, Jiří, 2003. *Interní audit a kontrola*. 2. přeprac. a dopl. vyd. Praha: C.H.Beck. C.H. Beck pro praxi. ISBN 80-7179-805-3.
22. EDWARDS, Kieran Jay., 2014. *Astronomy and big data: a data clustering approach to identifying uncertain galaxy morphology*. 1st edition. New York: Springer. ISBN 9783319065984.
23. FAWCETT, Tom, 2006. An introduction to ROC analysis. *Pattern recognition letters*. **27**(8), 861-874. ISSN 01678655.
24. FERRELL, O. C., John. FRAEDRICH a Linda. FERRELL, 2012. *Business ethics: ethical decision making and cases*. 9th ed. Boston: Houghton Mifflin Co. ISBN 9781111825164.
25. FISK, Marina Axelson, 2015. *Comparative Gene Finding: Models, Algorithms and Implementation*. Second Edition. London: Springer. ISBN 978-1-4471-6692-4.
26. Fraud, 2017. *Britannica Academic* [online]. [cit. 2017-10-20]. Dostupné z: http://academic.eb.com/levels/collegiate/article/fraud/35211
27. GEETHA, S. a Asnath Victy PHAMILA, 2016. *Combating security breaches and criminal activity in the digital sphere*. Hershey, PA: Information Science Reference, An Imprint of IGI Global. ISBN 9781522501930.
28. GIRISH, Palshikar Keshav, 2002. The hidden truth. *Intelligent enterprise (San Mateo, Calif.)*. **5**(9), 46-51. ISSN 15243621.
29. GIUSTI, Antonio., Gunter. RITTER a Maurizio VICHI, 2013. *Classification and data mining*. Berlin: Springer Verlag. Studies in classification, data analysis, and knowledge organization. ISBN 9783642288937.
30. GOLDMANN, Peter, 2009. *Anti-fraud risk and control workbook*. 1st. Hoboken, N.J.: John Wiley & Sons. ISBN 9780470496534.
31. GOLDMANN, Peter, 2010. *Fraud in the Markets: Why It Happens and How to Fight It*. 1st. Hoboken, New Jersey: Wiley. ISBN 978-0-470-50789-6.
32. GOTTSCHALK, Petter, 2010. Categories of financial crime. *Journal of financial crime*. **17**(4), 441-458. ISSN 13590790.
33. HALL, Mark A., 1999. *Correlation-based Feature Selection for Machine Learning*. The University of Waikato.
34. HALL, Mark A. a Lloyd A. SMITH, 1997. *Feature Subset Selection: A Correlation Based Filter Approach*.
35. HAYKIN, Simon, 1999. *Neural networks: a comprehensive foundation*. 2nd ed. Upper Saddle River (New Jersey): Prentice Hall International. ISBN 0139083855.
36. HOLLENBACK, Craig., 2007. *Industrial hygiene & safety auditing: a manual for practice*. 2nd ed. Fairfax, VA: AIHA Press. ISBN 9781931504829.
37. HOOGS, Bethany, 2007. A genetic algorithm approach to detecting temporal patterns indicative of financial statement fraud. *International journal of intelligent systems in accounting, finance & management*. **15**(1-2), 41-56. ISSN 1055615X.

38. HUANG, Shaio Yan, Chi-Chen LIN, An-An CHIU a David C. YEN, 2016. Fraud detection using fraud triangle risk factors. *Information systems frontiers*. 1-14. ISSN 13873326.
39. CHEN, Fu Hsiang, Der-Jang CHI a Jia-Yi ZHU, 2014b. Application of Random Forest, Rough Set Theory, Decision Tree and Neural Network to Detect Financial Statement Fraud – Taking Corporate Governance into Consideration. In: HUANG, D-S., ed., V. BEVILACQUA, ed. a P. PREMARATNE, ed. *International Conference on Intelligent Computing: ICIC 2014*. Cham: Springer, s. 221-234.
40. CHEN, Suduan, Yeong-Jia James GOO a Zone-De SHEN, 2014a. A Hybrid Approach of Stepwise Regression, Logistic Regression, Support Vector Machine, and Decision Tree for Forecasting Fraudulent Financial Statements. *The scientific world*. 1-9. ISSN 23566140.
41. CHOI, Stephen, Sara E. GILLEY a David F. MARCUS, 2016. *SEC Enforcement Activity Against Public Company Defendants: Fiscal Years 2010–2015* [online]. 16 s. [cit. 2017-10-20]. Dostupné z: https://www.cornerstone.com/Publications/Reports/SEC-Enforcement-Activity-Against-Public-Company-Defendants
42. *ISA 240: Postun auditorů při posuzování možných podvodů při auditu účetní uzávěrky* [online], 2004. 39 s. [cit. 2018-04-02]. Dostupné z: https://www.kacr.cz/data/Metodika/Auditing/ISA/ISA240.pdf
43. JAEGER, Jaclyn, 2011. Study: Fraud Reporting Hits Record Levels. *Compliance week*. London, **8**(94), 17-69. ISSN 1549957X.
44. JAIN, A. K., 1996. Artificial neural networks: a tutorial. *Computer (Long Beach, Calif.)*. **29**(3), 31-44. ISSN 00189162.
45. JANS, Mieke, Nadime LYBAERT a Koen VANHOF, 2010. *Data Mining and Economic Crime Risk Management In: Surveillance Technologies and Early Warning Systems: Data Mining Applications for Risk Detection*. 1st. Hershey, PA: Information Science Reference. ISBN 1616928654.
46. JANS, Mieke, Nadime LYBAERT a Koen VANHOOF MANAGEMENT ASSOCIATION, INFORMATION, ed., 2013. *Data Mining and Economic Crime risk Management* In: *Data Mining: Concepts, Methodologies, Tools, and Applications*. Hershey, PA: Information Science Reference. ISBN 978-1466624559.
47. JARVIS, Dennis, 2002. Zero tolerance. *Supply management*. **7**(1), 13. ISSN 13622021.
48. JOHNSON, Arthur T., 1985. Municipal Employee Assistance Programs: Managing Troubled Employees. *Public administration review*. **45**(3), 283-290. ISSN 00333352.
49. KASSEM, Rasha a Andrew HIGSON, 2012. The new fraud triangle model. *Journal of emerging trends in economics and management sciences*. **3**(3), 191-195. ISSN 21417024.
50. KOTSIANTIS, S., E. KOUMANAKOS, D. TZELEPIS a V. TAMPAKAS, 2006. Forecasting fraudulent financial statements using data mining. *International Journal of Computational Intelligence*. **3**(2), 104-110. ISSN 13042386.
51. KRAIPEERAPUN, Pawalai a Somkid AMORNSAMANKUL, 2015. Using stacked generalization and complementary neural networks to predict Parkinson's disease. *Natural Computation (ICNC), 2015 11th International Conference on*. IEEE, 1290-1294. ISSN 21579555.
52. KRANACHER, Mary-Jo, Richard RILEY a Joseph T. WELLS, 2011. *Forensic Accounting and Fraud Examination*. 1st. Hoboken, N.J.: John Wiley. ISBN 978-0-470-43774-2.
53. LAKSHMI, T. Miranda, 2013. An Analysis on Performance of Decision Tree Algorithms using Student's Qualitative Data. *International journal of modern education and computer science*. **5**(5), 18-27. ISSN 20750161.

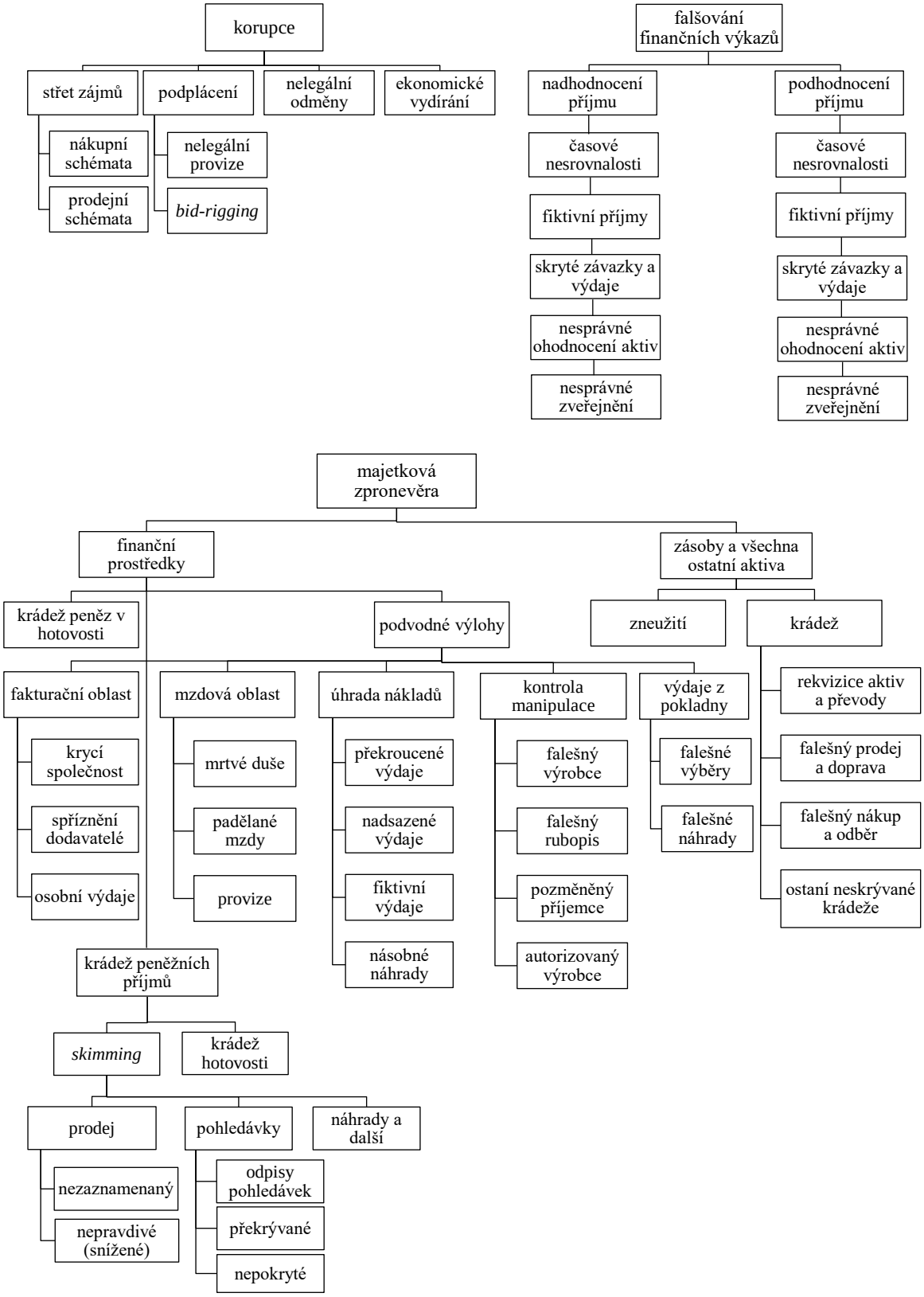
54. LENDEZ, Anthony M. a James J. KOREVEC, 1999. How to prevent and detect financial statement fraud. *The Journal of corporate accounting & finance*. **11**(1), 47-54. ISSN 10448136.
55. LI, Xinyang, Wei XU a Xuesong TIAN, 2014. How to protect investors? A GA-based DWD approach for financial statement fraud detection. *Systems, Man and Cybernetics, IEEE International Conference on*. IEEE, 3548-3554. ISSN 1062922X.
56. LOURIDAS, Panos a Christof EBERT, 2016. Machine Learning. *IEEE software*. **33**(5), 110-115. ISSN 07407459.
57. MADANI, Kurosh., 2011. *Computational intelligence*. Berlin: Springer. Studies in computational intelligence, v. 343. ISBN 9783642202056.
58. MARINI, Federico, 2007. Multilayer feed-forward artificial neural networks for class modeling. *Chemometrics and intelligent laboratory systems*. **88**(1), 118-124. ISSN 01697439.
59. MARSLAND, Stephen., 2015. *Machine learning: an algorithmic perspective*. Second edition. Boca Raton: CRC Press. ISBN 978-1-4665-8328-3.
60. MARTIN, Josh, 1998. An HR guide to white collar crime. *HR focus*. **75**(9), 1-14. ISSN 10596038.
61. MAŘÍK, Vladimír, Olga ŠTĚPÁNKOVÁ a Jiří LAŽANSKÝ, 2003. *Umělá inteligence*. Vyd. 1. Praha: Academia. ISBN 80-200-1044-0.
62. MENKUS, Belden, 1989. Point: The Future Of Internal Auditing Must Be Proactive. *The Internal auditor*. Altamonte Springs, **46**(4), 30-43. ISSN 00205745.
63. MITCHELL, Tom M., 1997. *Machine Learning*. New York: McGraw-Hill. ISBN 0070428077.
64. MOORE, Don A., Philip E. TETLOCK, Llozd TANTLU a Max H. BAYERMAN, 2006. Conflicts of interest and the case of auditor independence: Moral seduction and strategic issue cycling. *The Academy of Management review*. **31**(1), 10-29. ISSN 03637425.
65. MOORE, John, 2010. Preventing and Detecting Fraud, Waste, and Abuse. *The Public manager (Potomac, Md.)*. **39**(2), 72-74. ISSN 10617639.
66. MURPHY, Kevin P., 2012. *Machine learning: a probabilistic perspective*. Cambridge, MA: MIT Press. ISBN 978-0-262-01802-9.
67. NABHAN, Reem Abdul Latif a Nitham M. HINDI, 2009. Bank fraud: Perception of bankers in the state of qatar. *Academy of banking studies journal*. **8**(1-2), 15-38. ISSN 19392230.
68. NEAR, Janet P. a Marcia P. MICELI, 1985. Organizational dissidence: The case of whistle-blowing. *Journal of business ethics*. **4**(1), 1-16. ISSN 01674544.
69. NGAI, E.W.T., Yong HU, Y.H. WONG, Yijun CHEN a Xin SUN, 2011. The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision support systems and electronic commerce*. **50**(3), 559-569. ISSN 01679236.
70. NICOLAS, Patrick R., 2014. *Scala for Machine Learning*. Birmingham: Packt Publishing Ltd. ISBN 9781783558742.
71. O'GARA, John D., 2004. *Corporate Fraud: Case Studies in Detection and Prevention*. 1st. New Jersey: Wiley. ISBN 978-0471493501.
72. OLEJ, Vladimír a Petr HÁJEK, 2010. *Úvod do umělé inteligence: moderní přístupy : distanční opora*. Vyd. 1. Pardubice: Univerzita Pardubice. ISBN 9788073953072.

73. PAI, Ping-Feng, Ming-Fu HSU a Ming-Chieh WANG, 2011. A support vector machine-based model for detecting top management fraud. *Knowledge-based systems*. **24**(2), 314-321. ISSN 09507051.
74. PERSONS, Obeua S., 2011. Using Financial Statement Data To Identify Factors Associated With Fraudulent Financial Reporting. *Journal of applied business research*. **11**(3), 38-46. ISSN 08927626.
75. QUI, Chen, 2017. Randomly selected decision tree for test-cost sensitive learning. *Applied soft computing*. **53**, 27-33. ISSN 15684946.
76. RAVISANKAR, P., V. RAVI, G. RAO RAGHAVA a I. BOSE, 2011. Detection of financial statement fraud and feature selection using data mining techniques. *Decision support systems and electronic commerce*. **50**(2), 491-500. ISSN 01679236.
77. ROKACH, Lior a Oded MAIMON, 2005. Top-Down Induction of Decision Trees Classifiers—A Survey. *IEEE transactions on systems, man and cybernetics. Part C, Applications and reviews*. **35**(4), 476-487. ISSN 10946977.
78. RUSSELL, Stuart J., Peter NORVIG a Ernest DAVIS, 2010. *Artificial intelligence: a modern approach*. 3rd ed. Boston: Pearson. Prentice Hall series in artificial intelligence. ISBN 978-0-13-207148-2.
79. SEETHARAMAN, A., M. SENTHILVELMURUGAN a Rajan PEIYANAYAGAN, 2004. Anatomy of computer accounting frauds. *Managerial auditing journal*. Bradford, **19**(8), 1055-1072. ISSN 02686902.
80. SCHUCHTER, Alexander a Michael LEVI, 2016. The Fraud Triangle revisited. *Security journal*. Basingstoke, **29**(2), 107-121. ISSN 09551662.
81. SCHULD, Maria, Ilya SYNAYSKI a Francesco PETRUCCIONE, 2015. An introduction to quantum machine learning. *Contemporary Physics*. **56**(2), 1-19. ISSN 172185.
82. SILVA, Catarina. a Bernardete. RIBEIRO, 2010. *Inductive inference for large scale text classification: kernel approaches and techniques*. Berlin: Springer. Studies in computational intelligence, v. 255. ISBN 9783642045325.
83. SINGLETON, Tomie W., Aaron J. SINGLETON, G. Jack BOLOGNA a Robert J. LINDQUIST, 2006. *Fraud auditing and forensic accounting*. 3rd ed. Hoboken, New Jersey: John Wiley. ISBN 978-0-470-05372-0.
84. SINNETT, William M., 2003. Ask FERF (financial executives research foundation) about... managing business risk. *Financial executive (1987)*. Morristown, **19**(4), 63. ISSN 08954186.
85. SONG, Xin-Ping, Zhi-Hua HU, Jian-Guo DU a Zhao-Han SHENG, 2014. Application of Machine Learning Methods to Risk Assessment of Financial Statement Fraud: Evidence from China Risk Assessment of Financial Statement Fraud. *Journal of forecasting*. **33**(8), 611-626. ISSN 02776693.
86. TRENERRY, Alan, 1999. *Principles of internal control*. 1st. Sydney, N.S.W: UNSW Press. ISBN 0-86840-401-2.
87. TURNER, Leslie a Andrea WEICKGENANNT, 2009. *Accounting Information Systems: Controls and Processes*. 1st. Hoboken: Wiley. ISBN 13-9780471479512.
88. VALLABHANENI, S. Rao., 2005. *Wiley CIA exam review*. 3rd ed. Hoboken, N.J.: Wiley. ISBN 0471718823.
89. WALLACE, Manolis, Ioannis E. ANAGNOSTOPOULOS, Phivos MYLONAS a Maria BIELIKOVA, 2010. *Semantics in Adaptive and Personalized Services: Methods, Tools and Applications*. 1. Berlin: Springer. ISBN 9783642116834.
90. WELYTOK, Jill Gilbert., 2006. *Sarbanes-Oxley for dummies*. 1st. Hoboken, NJ: Wiley. ISBN 9780471768463.

91. WEST, Jarrod a Maumita BHATTACHARYA, 2015. Mining Financial Statement Fraud: An Analysis of Some Experimental Issues. *Industrial Electronics and Applications (ICIEA), 2015 IEEE 10th Conference on*. IEEE, 461-466.
92. WEST, Jarrod a Maumita BHATTACHARYA, 2016. Intelligent financial fraud detection: a comprehensive review. *Computers & security*. **57**, 47-66. ISSN 01674048.
93. What We Do, 2013. *U.S. Securities and Exchange Commission* [online]. [cit. 2017-11-15]. Dostupné z: https://www.sec.gov/Article/whatwedo.html
94. WIERSEMA, William, 2015. Preventing financial fraud. *Electrical apparatus*. **68**(5), 40-41. ISSN 01901370.
95. WILLIAMS, Nigel, Sebastian ZANDER a Grenville ARMITAGE, 2006. A preliminary performance comparison of five machine learning algorithms for practical IP traffic flow classification. *ACM SIGCOMM Computer Communication Review*. **36**(5), 5-16.
96. WITTEN, I. H., Eibe. FRANK a Mark A. HALL, 2011. *Data mining: practical machine learning tools and techniques*. 3rd ed. Burlington, MA: Morgan Kaufmann. Morgan Kaufmann series in data management systems. ISBN 978-0-12-374856-0.
97. WULF, Katharina, 2011. *From codes of conduct to ethics and compliance programs recent developments in the United States*. 1st. Berlin: Logos-Verl. ISBN 9783832528980.
98. YOUNG, Michael R., 2014. *Financial Fraud Prevention and Detection: Governance and Effective Practices*. 1st. Hoboken, New Jersey: Wiley. ISBN 978-1-118-61763-2.
99. YUNIARTI, Rozmita Dewi, 2017. The effect of internal control and anti-fraud awareness on fraud prevention: (A survey on inter-governmental organizations). *Journal of Economics, Business & Accountancy*. Indonesie, **20**(1), 113-124. ISSN 2088785X.
100. ZACK, Gerard M., 2009. *Fair value accounting fraud: new global risks and detection techniques*. 1st. Hoboken, N.J.: John Wiley. ISBN 978-0-470-47858-5.

SEZNAM PŘÍLOH

Příloha A



Obrázek 19: Klasifikace interních podvodů dle ACFE

Zdroj: (ACFE, 2016, s. 11)

Příloha B

Tabulka 11: Popisné statistiky

finanční proměnné	jednotky	průměr	minimum	maximum	sm. odchylka
růst prodeje	procento	4,218E+01	-5,259E+01	1,774E+03	1,502E+02
růst pohledávek	procento	3,765E+01	-9,138E+01	3,199E+02	6,243E+01
růst čistého provozního toku z provozní činnosti	procento	-5,554E+01	-3,195E+04	1,216E+04	2,773E+03
závazky	měna [USD]	1,222E+11	1,010E+02	7,230E+12	6,471E+11
aktiva	měna [USD]	1,575E+11	8,058E+03	1,110E+13	9,364E+11
hrubý zisk	měna [USD]	2,153E+10	-6,500E+07	2,490E+12	1,971E+11
čistý příjem	měna [USD]	3,751E+09	-2,290E+09	4,208E+11	3,333E+10
prodej	měna [USD]	7,048E+10	1,010E+02	9,670E+12	7,640E+11
hotovost a hotovostní ekvivalenty	měna [USD]	1,552E+10	1,010E+02	5,607E+11	7,283E+10
pohledávky	měna [USD]	4,817E+10	1,010E+02	2,890E+12	2,608E+11
peněžní tok (peněžní činnost)	měna [USD]	4,190E+09	-4,800E+08	3,068E+11	2,539E+10
peněžní tok (investiční činnost)	měna [USD]	-5,599E+09	-5,502E+11	3,708E+10	4,492E+10
peněžní tok (finanční činnost)	měna [USD]	1,632E+09	-9,544E+10	2,288E+11	2,138E+10
oběžná aktiva	měna [USD]	9,408E+10	1,010E+02	5,620E+12	5,062E+11
pracovní kapitál	měna [USD]	5,905E+09	-2,707E+11	1,320E+12	1,095E+11
pracovní kapitál/aktiva		1,617E+00	-1,594E+02	1,010E+02	2,035E+01
zásoby/prodej		2,914E+01	1,727E-03	1,010E+02	4,579E+01
zásoby/aktiva		2,912E+01	6,390E-04	1,010E+02	4,581E+01
hrubý příjem/aktiva		2,192E+00	-1,853E+00	1,010E+02	1,371E+01
čistý zisk/aktiva		-1,201E+00	-1,899E+02	1,280E+00	1,502E+01
oběžná aktiva/aktiva		3,016E+00	1,541E-02	1,010E+02	1,574E+01
čistý zisk/prodej		1,520E+00	-2,893E+01	1,010E+02	1,419E+01
pohledávky/prodej		7,355E+00	4,878E-04	1,010E+02	2,325E+01
prodej/aktiva		2,776E+00	1,803E-02	1,010E+02	1,365E+01
oběžná aktiva/krátkodobé závazky		4,938E+00	4,294E-03	1,010E+02	1,629E+01
hotovost a hotovostní ekvivalenty/aktiva		5,802E+00	1,202E-03	1,010E+02	2,331E+01
zásoby/krátkodobé závazky		3,003E+01	8,868E-04	1,010E+02	4,592E+01
čistý zisk/hrubý příjem		1,302E+01	-1,690E+01	1,751E+03	1,389E+02
závazky/aktiva		4,260E+00	1,715E-02	2,681E+02	2,504E+01
hotovost/oběžná aktiva		7,200E+00	2,414E-03	1,010E+02	2,557E+01
pohledávky/aktiva		5,248E+00	2,509E-04	1,010E+02	2,204E+01
hrubý zisk/prodej		3,541E+00	-1,350E+01	1,010E+02	1,760E+01
dlouhodobý majetek/aktiva		7,143E+00	2,036E-10	1,010E+02	2,558E+01
pohledávky/oběžná aktiva		6,700E+00	6,875E-04	1,010E+02	2,443E+01
dlouhodobý dluh/aktiva		1,769E+01	6,776E-05	1,010E+02	3,813E+01

Zdroj: vlastní zpracování (Statistica 12)

Příloha C

Tabulka 12: Ukázka sběru dat

růst prodeje	**růst pohledávek**	růst čistého provozního toku z provozní činnosti	závazky	aktiva	**hrubý zisk**	**čistý příjem**	prodej	**peněžní prostředky a peněžní ekvivalenty**	pohledávky	peněžní tok (peněžní činnosti)	peněžní tok (investiční činnosti)	peněžní tok (finanční činnosti)	**oběžná aktiva**	pracovní kapitál	pracovní kapitál/aktiva	**čistý zisk/prodej**	zásoby/aktiva	hrubý příjem/aktiva
-0,56	?	-37,57	23820000	47810000	3170000	1120000	5520000	10370000	80000	-470000	-900000	9620000	10450000	9660000	0,2020	0,2029	?	0,0663
?	?	?	198250000	721770000	116550000	5090000	221790000	23510000	51190000	11500000	-176360000	142340000	159240000	103820000	0,1438	0,0229	0,0773	0,1615
?	?	-6	308680000000	343660000000	20560000000	4850000000	21400000000	13960000000	205080000000	7310000000	-30200000000	22360000000	?	?	?	0,2266	0,0208	0,0598
4,7	20,09	125,89	15760000	42790000	19970000	-460000	96840000	1430000	13400000	1100000	-3640000	420000	33330000	17970000	0,4200	-0,0048	?	0,4667
7,64	10,6	-31,79	361520000	875200000	206710000	41130000	784680000	2970000	186620000	111370000	-34810000	75390000	221010000	64690000	0,0739	0,0524	?	0,2362
-1,63	-6,59	-19,07	48190000000	105130000000	31710000000	6260000000	48130000000	13530000000	8540000000	12380000000	-2890000000	6900000000	33180000000	16930000000	0,1610	0,1301	0,0595	0,3016
?	?	?	54910000000	121270000000	56220000000	16980000000	73750000000	6600000000	15780000000	31630000000	-24790000000	-9410000000	85080000000	52390000000	0,4320	0,2302	0,0094	0,4636
-0,39	?	-22,62	5880000000	6410000000	1290000000	72730000	3210000000	?	401060000	304890000	-116600000	-114300000	?	?	?	0,0227	0,0088	0,2012
11,96	?	-25,53	19840000	90600000	68900000	14690000	68910000	5260000	7700000	28830000	-2240000	-12780000	72480000	54960000	0,6066	0,2132	?	0,7605
4,55	63,89	9,18	4530000	9470000	9630000	360000	27810000	1820000	2520000	750000	-920000	-40000	4670000	480000	0,0507	0,0129	?	1,0169
29,02	11,81	5,35	3140000000	6700000000	1670000000	260900000	4580000000	209300000	1170000000	632000000	-719000000	16400000	2570000000	730000000	0,1090	0,0570	0,1327	0,2493
-16,08	319,86	55,97	47660000	26900000	11370000	18180000	13200000	10370000	2100000	-6690000	-940000	9760000	13770000	-25230000	-0,9379	1,3773	0,0152	0,4227
6,38	4,25	112,82	2460000000	4400000000	661000000	180000000	3300000000	120000000	687000000	415000000	-1180000000	606000000	1440000000	521000000	0,1184	0,0545	0,1211	0,1502
-10,38	10,36	-18,97	1780000000	2970000000	99590000	580000	614530000	79450000	112010000	388440000	-637000000	270680000	228340000	15210000	0,0051	0,0009	?	0,0335
6,55	0,58	15,74	1110000000	3100000000	3470000000	1040000000	4760000000	917220000	174460000	1220000000	-259430000	-741880000	1800000000	1081840000	0,3490	0,2185	0,1627	1,1194
4,38	?	-32,39	136780000000	155670000000	33330000000	4430000000	33810000000	2020000000	110220000000	7080000000	-6550000000	-3270000000	132470000000	74750000000	0,4802	0,1310	?	0,2141
13,16	8,63	60,06	2070000000	2060000000	705990000	312490000	873590000	130740000	26530000	537630000	-1440000000	-277750000	1640000000	944980000	0,4587	0,3577	?	0,3427
-3,93	7,48	-26,1	5570000000	12550000000	4540000000	1650000000	17340000000	1370000000	2720000000	1530000000	-982000000	-694000000	7170000000	4030000000	0,3211	0,0952	0,1769	0,3618
4,81	10,51	8,84	3850000000	6590000000	7720000000	856900000	9720000000	1350000000	1060000000	1130000000	-428300000	-585100000	3860000000	1730000000	0,2625	0,0882	0,1493	1,1715
?	?	-280,61	534210000	583670000	22630000	2910000	25330000	27570000	313450000	-14910000	-31470000	57780000	365790000	-115160000	-0,1973	0,1149	0,0424	0,0388
-22,41	-32,85	-108,75	350820000	447910000	140920000	16740000	342540000	95220000	52280000	65670000	-38310000	-12870000	257140000	10560000	0,0236	0,0489	0,1744	0,3146
43,84	17,82	5,34	59500000	235330000	28110000	12690000	93720000	28090000	6690000	35420000	-23760000	30000	62040000	50250000	0,2135	0,1354	0,1127	0,1194
?	?	-59,35	1680000000000	1870000000000	83100000000	7520000000	90710000000	399900000000	652500000000	16070000000	16780000000	-25360000000	1373330000000	6940000000	0,0037	0,0829	?	0,0444
?	?	-81,96	1590000000	1710000000	64780000	15190000	68330000	77220000	1070000000	5660000	-58570000	92320000	1150000000	-410000000	-0,2398	0,2223	?	0,0379
-12,2	-9,23	-1,95	5230000000	5690000000	2890000000	114990000	11460000000	78980000	881410000	440540000	-174180000	-302460000	2200000000	860000000	0,1511	0,0100	0,0717	0,5079
-8,97	14,37	5,65	2240000000	3520000000	1440000000	106300000	3800000000	212000000	523600000	413100000	-294300000	-263400000	1920000000	480000000	0,1364	0,0280	0,0788	0,4091
1,64	1,7	3,38	7230000000	11360000000	3950000000	1110000000	7710000000	1670000000	1250000000	1690000000	-833250000	-370170000	5320000000	3340000000	0,2940	0,1440	0,1092	0,3477
8,78	-4,01	130,04	5970000000	10700000000	2010000000	743000000	32960000000	1640000000	1220000000	1590000000	-696000000	-150000000	4640000000	1760000000	0,1645	0,0225	0,1477	0,1879
?	?	-56,57	152460000000	173440000000	10810000000	1960000000	11240000000	8260000000	119300000000	2010000000	4740000000	-3010000000	137000000000	-1970000000	-0,0114	0,1744	0,0196	0,0623
3,83	-25,71	-116,99	15990000000	27280000000	3290000000	406000000	60990000000	570000000	2940000000	-444000000	-980000000	1210000000	17130000000	5570000000	0,2042	0,0067	0,2559	0,1206
-10,74	-20,5	-7,32	5650000000	11210000000	2630000000	1620000000	5950000000	1670000000	597780000	1860000000	-1740000000	-865330000	3420000000	2636420000	0,2352	0,2723	0,0619	0,2346
25,46	39,59	-36,88	120900000	239940000	59450000	-1800000	691450000	7100000	65370000	13710000	-7750000	-5050000	109730000	-3250000	-0,0135	-0,0026	0,1329	0,2478
7,57	24,07	0,09	3420000000	2700000000	835500000	157700000	3040000000	280900000	129400000	346900000	-258700000	83200000	465100000	-41300000	-0,0153	0,0519	0,0070	0,3094
34,2	?	-40,66	1020000000	1300000000	220250000	13780000	572540000	24230000	439660000	-162320000	181690000	-19090000	471060000	-475520000	-0,3658	0,0241	?	0,1694
-2,85	36,47	-26,73	15990000	36290000	4670000	1780000	11380000	2180000	3380000	3870000	-12420000	3950000	6380000	780000	0,0215	0,1564	0,0102	0,1287
-3,94	-31,54	-6,59	5740000000	9240000000	3860000000	1840000000	14350000000	1810000000	2000000000	3050000000	-825000000	-2220000000	5410000000	2800000000	0,3030	0,1282	0,0924	0,4177
39,94	56,64	37,26	98030000	277120000	258610000	111270000	688280000	55060000	67380000	119670000	-54380000	-41200000	157630000	69460000	0,2506	0,1617	0,0839	0,9332
32,4	111,26	-61,41	600960000	1010000000	75450000	-7850000	320240000	272860000	131550000	10650000	-245440000	403420000	595360000	420040000	0,4159	-0,0245	0,0016	0,0747
8,76	-57,65	-115,13	13820000	14820000	810000	1150000	2570000	4880000	1680000	-2910000	5230000	1070000	7580000	-3900000	-0,2632	0,4475	0,0209	0,0547

Zdroj: vlastní zpracování

Příloha D

Tabulka 13: Vyzkoušené kombinace nastavení parametrů k jednotlivým metodám

klasifikátor	nastavení
MLP	trénovacíČas = {500, 1000}, skrytáVrstva = {1, 5, 7, 9}
	momentum = {0.2, 0.4, 0.6, 0.8, 1}, rychlostUčení = {0.2, 0.4, 0.6, 0.8, 1}
C4.5	faktorSpolehlivosti = {0.25, 0.35, 0.4, 0.5}, minPočetInstancíNaList: 10-60 (krok 5)
SMO	parametrKomplexnosti: 1-60 (krok 5)
	jádrováFunkce = {polyKernel (exponent: 1, 2), RBFkernel (gamma: 0.001, 0.2), Puk, NormalizedPolyKenel
Bagging	klasifikátor: MLP, SMO, C4.5
	početIterací = {10, 50, 100}
	velikostBaguJakoProc.Tren.Dat: 10-100 (krok 5)
Boosting	klasifikátor: MLP, SMO, C4.5
	MLP: trénovacíČas = 500, rychlostUčení = {0.1, 0.2, 0.3}, skrytáVrstva: 1-50 (krok 5)
	SMO: parametrKomplexnosti: 1-60 (krok 5), jádrováFunkce = {polyKernel (exponent: 1, 2), RBFkernel (gamma: 0.001, 0.2), Puk, NormalizedPolyKernel}
	C4.5: faktorSpolehlivosti = {0.25, 0.35, 0.4, 0.5}, minPočetInstancíNaList: 10-50 (krok 5)
Stacking	klasifikátor = {MLP, C4.5}, metaKlasifikátor: SMO
	jádrováFunkce = {polyKernel (exponent: 1, 2), RBFkernel (gamma: 0.001, 0.2)
	klasifikátor = {SMO, MLP}, metaKlasifikátor: C4.5
	faktorSpolehlivosti = 0.25, minPočetInstancíNaList: 10-60 (krok 5)
	klasifikátor = {SMO, C4.5}, metaKlasifikátor: MLP
	trénovacíČas = 500, rychlostUčení = {0.1, 0.2, 0.3}, skrytáVrstva: 1-50 (krok 5)

Zdroj: vlastní zpracování

Příloha E

Tabulka 14: Podrobnější výsledky měření (MLP)

číslo běhu	přesnost (%)	senzitivita	specificita	plocha pod ROC
1	51,25	0,300	0,725	0,537
2	56,25	0,875	0,250	0,713
3	56,25	0,875	0,250	0,629
4	56,25	0,825	0,300	0,528
5	55,00	0,875	0,225	0,575
6	57,50	0,750	0,400	0,602
7	65,00	0,650	0,650	0,655
8	53,75	0,400	0,675	0,571
9	65,00	0,975	0,325	0,653
10	58,75	0,425	0,750	0,598
průměr	57,50	0,695	0,455	0,606

Zdroj: vlastní zpracování

Tabulka 15: Podrobnější výsledky měření (C4.5)

číslo běhu	přesnost (%)	senzitivita	specificita	plocha pod ROC
1	53,75	0,675	0,400	0,547
2	50,00	1,000	0,000	0,500
3	50,00	1,000	0,000	0,500
4	57,50	0,675	0,475	0,598
5	50,00	1,000	0,000	0,500
6	57,50	0,650	0,500	0,600
7	50,00	1,000	0,000	0,500
8	53,75	0,650	0,425	0,570
9	60,00	0,750	0,450	0,630
10	62,50	0,700	0,550	0,651
průměr	54,50	0,810	0,280	0,560

Zdroj: vlastní zpracování

Tabulka 16: Podrobnější výsledky měření (SMO)

číslo běhu	přesnost (%)	senzitivita	specificita	plocha pod ROC
1	53,75	0,850	0,225	0,538
2	53,75	1,000	0,075	0,538
3	51,25	0,925	0,100	0,513
4	58,75	0,900	0,275	0,588
5	61,25	0,900	0,325	0,613
6	60,00	0,975	0,225	0,600
7	70,00	0,825	0,575	0,700
8	60,00	0,925	0,275	0,600
9	55,00	0,925	0,175	0,550
10	56,25	0,875	0,250	0,563
průměr	58,00	0,910	0,250	0,580

Zdroj: vlastní zpracování

Tabulka 17: Podrobnější výsledky měření (Bagging)

číslo běhu	přesnost (%)	senzitivita	specificita	plocha pod ROC
1	53,75	0,550	0,525	0,559
2	63,75	0,600	0,675	0,641
3	63,75	0,750	0,525	0,649
4	58,75	0,675	0,500	0,576
5	65,00	0,700	0,600	0,631
6	51,25	0,550	0,475	0,547
7	57,50	0,575	0,575	0,616
8	56,25	0,550	0,575	0,583
9	58,75	0,600	0,575	0,626
10	58,75	0,650	0,525	0,657
průměr	58,75	0,620	0,555	0,608

Zdroj: vlastní zpracování

Tabulka 18: Podrobnější výsledky měření (Boosting)

číslo běhu	přesnost (%)	senzitivita	specificita	plocha pod ROC
1	51,25	0,800	0,225	0,540
2	60,00	0,825	0,375	0,580
3	53,75	0,950	0,125	0,548
4	53,75	0,575	0,500	0,533
5	67,50	0,800	0,550	0,650
6	58,75	0,975	0,200	0,634
7	60,00	0,875	0,325	0,579
8	60,00	0,500	0,700	0,629
9	57,50	0,775	0,375	0,582
10	57,50	0,875	0,275	0,579
průměr	58,00	0,795	0,365	0,585

Zdroj: vlastní zpracování

Tabulka 19: Podrobnější výsledky měření (Stacking)

číslo běhu	přesnost (%)	senzitivita	specificita	plocha pod ROC
1	56,25	0,475	0,650	0,563
2	53,75	0,900	0,175	0,538
3	56,25	0,925	0,200	0,563
4	55,00	0,825	0,275	0,550
5	57,50	0,900	0,250	0,575
6	53,75	0,150	0,925	0,538
7	55,00	0,925	0,175	0,550
8	57,50	0,625	0,525	0,575
9	48,75	0,850	0,125	0,488
10	53,75	0,925	0,150	0,538
průměr	54,75	0,750	0,345	0,548

Zdroj: vlastní zpracování