

Univerzita Hradec Králové
Fakulta informatiky a managementu
Katedra ekonomie

Metody Business Intelligence ve finanční firmě
Diplomová práce

Autor: Bc. Vojtěch Boukal

Studijní obor: Informační management (2)

Vedoucí práce: Ing. Eva Hamplová, Ph.D.

Odborný konzultant: Ing. Filip Mezera, AKCENTA CZ, a.s.

Hradec Králové

Srpen 2016

Prohlášení:

Prohlašuji, že jsem diplomovou práci zpracoval samostatně a s použitím uvedené literatury.

V Hradci Králové dne 19.8.2016

Vojtěch Boukal

Poděkování

Rád bych poděkoval vedoucí práce Ing. Evě Hamplové, Ph.D. za cenné připomínky a rady, a také konzultantovi Ing. Filipu Mezerovi z firmy AKCENTA CZ, a.s. za poskytnutí dat a odborný dohled nad výzkumnou částí práce.

Anotace

Diplomová práce se zabývá analýzou klientů finanční společnosti AKCENTA CZ, a.s. Požadavkem společnosti bylo odhalení bonitních klientů na počátku smluvního vztahu a odhalení podezřelých či rizikových klientů. Cíle bylo dosaženo za pomoci rozboru dat a klasifikačních data miningových metod. V diplomové práci je využit software SPSS Modeler, který slouží k podrobnějšímu zkoumání dat a k sestavení konečného modelu. V práci je dále využita metodika CRISP-DM, která popisuje doporučený postup od pochopení dat, až po samotnou implementaci výsledků. Výstupem modelu je klasifikační strom, který predikuje budoucí profitabilitu klientů. V závěru práce jsou shrnuty výsledky a navržena doporučení, které budou následně sloužit finanční společnosti pro lepší rozhodování.

Annotation

Title: Methods of business intelligence in financial company

The diploma thesis deals with the analysis of financial company AKCENTA CZ a.s. clients. The aim of the company was revealing creditworthy clients at the beginning of the contractual relationship and revelations of suspicious or high-risk clients. The objectives were achieved with the help of data analysis and classification of data mining methods. SPSS Modeler software is used for detailed examination of the data and compilation of the final model. In the thesis methodology CRISP-DM is applied, which describes recommended process of understanding the data to the actual implementation of the results. The output of the model is a classification tree that predicts future profitability of clients. In the conclusion the results are summarized and recommendations, which will then be used by the financial company to make better decisions, are proposed.

Obsah

1 Úvod	1
2 Cíl a metodika práce.....	2
3 Teoretická východiska	3
3.1 Business intelligence	3
3.1.1 Business intelligence (BI) a data mining (DM).....	4
3.1.2 Metody data miningu	6
3.1.3 Proces data miningu.....	8
3.2 Finanční instituce.....	13
3.2.1 Funkce finančních institucí na finančním trhu	14
3.2.2 Začlenění společnosti Akcenta CZ, a.s. na finanční trh.....	20
4 Aplikace BI nástrojů ve finanční instituci.....	23
4.1 Datový soubor	23
4.2 Rozdělení klientů	25
4.2.1 První rozdělení klientů.....	27
4.2.2 Druhé rozdělení klientů	27
4.2.3 Konečné rozdělení klientů	28
4.3 Vytvoření nových proměnných.....	30
4.3.1 Označení jednorázových (one-time) klientů.....	30
4.3.2 Sumarizace objemu a profitu	31
4.3.3 Průměrná doba mezi transakcemi.....	31
4.3.4 Nejmenší a největší transakce	31
4.3.5 Nejmenší a největší profit z transakcí	32
4.3.6 Průměrný profit a objem	32
4.3.7 Medián profitu a objemu.....	32
4.4 Aplikace BI nástrojů.....	32

4.4.1	Analýza klíčových proměnných	33
4.4.2	Právní forma klientů.....	36
4.4.3	One-time klienti.....	38
4.5	Odhalení bonitního klienta na začátku smluvního vztahu.....	44
5	Shrnutí výsledků.....	48
6	Závěr.....	50
7	Seznam použité literatury.....	51
8	Seznam obrázků, grafů a tabulek.....	53

1 Úvod

V posledních letech s rozmachem informatiky a počítačových věd dochází k vytváření čím dál většího množství dat. Podle odhadu společnosti IDC (International Data Corporation) by mělo množství dat v roce 2020 dosáhnout 44 ZB (trilionu gigabitů). Problémem tak ve většině případů není jejich nedostatek, ale spíše nalezení způsobu, jak získat potřebné informace a ty následně aplikovat k maximalizaci užitku. (EMC Corporation Online, 2014)

Jedním z předních oborů, který se tímto problémem zabývá je i business intelligence. Jeho použití je už běžnou součástí velkých společností, kde hraje významnou roli především v oblasti usnadňující rozhodování a předpovídání. Čím dál více se však tento obor dostává i do povědomí menších podniků, které ho mohou využít jako konkurenční výhodu.

Pro diplomovou práci byla použita reálná data od finanční instituce AKCENTA CZ, a.s. čítající informace o 2 594 klientech a jejich uskutečněných transakcích. V práci jsou poté popsány praktiky a postup, jak z těchto dat získat relevantní informace využitelné pro společnost.

Práce je rozdělena do dvou dílčích celků. Teoretická část, se zaměřuje na seznámení s metodami business intelligence a data mining nástrojů. Součástí teoretické části je také kapitola, která se věnuje porovnáním a funkcím finančních institucí na území České republiky. Teoretické poznatky a zjištění jsou následně aplikovány v druhé, praktické části práce.

2 Cíl a metodika práce

Cílem diplomové práce je odhalení bonitního klienta v počáteční fázi smluvního vztahu prostřednictvím metod business intelligence. Společnost by tak následně měla být schopna rozpoznat, zda pro ni klient bude ziskový či nikoliv a na základě toho přizpůsobit zařazení klienta ke vhodnému sazebníku. Nevýhodné klienty pak bude schopna odradit od spolupráce úplně, či ztrátu vykompenzovat vyššími poplatky, naopak ziskové klienty je nutné udržet a motivovat k dlouhodobému využívání finančních služeb.

Data jsou rozdělena do tří datových sad. První datová sada slouží především k identifikaci klienta. Kromě ID klienta je k dispozici jeho právní forma podnikání (právnícká, fyzická podnikatel, fyzická ne podnikatel), údaje o objemu, profitu a počtu transakcí. Data jsou navíc rozdělena na krátkodobá (první 3 měsíce), střednědobá (prvních 6 měsíců), dlouhodobá (6-12 měsíců) a kumulovaná (1-12 měsíců). Druhá datová sada poskytuje informace o prvních pěti uskutečněných transakcích (doba mezi transakcemi, profitem, objemem) a poslední sada je oproti druhé omezena na tři měsíce od podpisu smlouvy.

V práci je rozebírána metodika rozdělení klientů podle jejich bonity a rentability v dlouhém a krátkém období. K rozdělení je použit tabulkový nástroj Microsoft Excel, který je dále nápomocen i pro přípravu dat a vytvoření nových proměnných. Pro samotnou aplikaci business intelligence a data mining nástrojů je využit statistický software od společnosti IBM, SPSS Modeler.

Dílními cíli je hledání odlišných znaků klientů v poskytnutých datech, odhalení rizikových a podezřelých klientů.

Použitými metodikami je analýza a sestavení modelů, tedy rozčlenění dat do menších celků, jejich následné zkoumání a sledování souvislostí mezi nimi. Současně s analýzou je za pomoci business intelligence a statistického softwaru IBM SPSS Modeler sestaven model, který slouží jako prostředek pro zjednodušení vnímání celé problematiky a je schopen pomoci najít společné znaky mezi klienty.

3 Teoretická východiska

Následující kapitola se zabývá teoretickými znalostmi, které budou následně sloužit ke splnění cílů diplomové práce. Teoretická část je rozdělena na dvě části. První z nich je business intelligence, kde jsou rozebírány metody zkoumání a porozumění dat. Druhá část se zabývá finančními institucemi, jejich rozdělením a popsáním hlavních rozdílů mezi nimi.

3.1 Business intelligence

Dnešní počítačová doba se vyznačuje obrovským množstvím dat a informací, které jsou vytvořeny a následně zaznamenávány. Podle již v úvodu zmíněné studie provedené společností IDC se počet digitálních informací zdvojnásobí každé dva roky. V roce 2013 je odhadováno množství dat vytvořených člověkem na 4,4 ZB (44 trilionu gigabytu), v roce 2020 by to mělo být až 44 ZB Pro představu studie uvádí, že kdyby data byla uložena v dotykových tabletech (iPad) s kapacitou 128 GB a ty by byly naskládány na sebe, v roce 2013 by dosáhly vzdálenosti dvou třetin od země na měsíc. V roce 2020 by tuto vzdálenost naplnily 6,6 krát. (EMC Corporation Online, 2014)

Obrázek č. 1: Množství dat vhodných k dalšímu zpracování



Zdroj: (EMC Corporation Online, 2014)

Zajímavým faktem je, že v roce 2013 bylo pro další analýzu použitelných pouze 22 % dat a pouze 5 % dat bylo skutečně analyzováno. Je to především z důvodu neznámého původu dat a jejich špatné strukturalizace. V roce 2020 by se podle odhadu mělo množství dat vhodných k dalšímu zpracování navýšit na více než 35 % především z důvodu automatického zpracování dat. (EMC Corporation Online, 2014)

Z trendu stále přibývajících množství dat je čím dál větší důraz kladen na rozlišení relevantních informací, které mohou být využity v náš prospěch. Právě to je hlavním cílem business intelligence.

Pod pojmem business intelligence se můžeme setkat s velkým množstvím různých definic:

-BI je vědomá, soustavná transformace dat ze všech dostupných zdrojů do nových forem, poskytujících informace, která je řízena firemní činností a orientována na výsledky. (Bierre, 2003, str. 18)

-BI je sada technik a nástrojů pro získávání a transformování nezpracovaných dat do smysluplných a užitečných informací pro analytické a podnikatelské potřeby. (Turner, 2016, online)

Business intelligence tak jednoduše řečeno můžeme chápat jako souvislý způsob získávání informací z dat, která následně pomáhají k dosahování podnikatelských cílů.

3.1.1 Business intelligence (BI) a data mining (DM)

V souvislosti s business intelligence se můžeme setkat také s pojmem **data mining** (dolování dat), který je nyní chápán jako součást procesu vyhledávání znalostí z dat. Původně se pojem data mining používal v návaznosti na statistickou exploraci dat a až později se tento termín začal vyskytovat v kontextu s intenzivní analýzou velkého rozsahu dat.

Pro aplikační oblasti, spojené zejména s podnikatelskou sférou, se proces data mining a vyhledávání znalostí z dat označuje business intelligence (BI). Proces je navržený a implementovaný tak, aby jeho výsledky byly interpretovatelné a přímo využitelné pro podnikatelské rozhodování. Tyto směry využití se někdy označují DM-BI. (Skalská, 2010, str. 14)

Z tohoto popisu lze pochopit úzké spojení mezi data mining a business intelligence.

Dolování dat umožňuje pomocí speciálních algoritmů objevovat v datech strategické informace. Je to analytická technika pevně spjatá s datovými sklady jako velmi kvalitním datovým zdrojem pro tyto speciální analýzy. Dolování dat lze charakterizovat jako proces extrakce relevantních, předem neznámých nebo nedefinovaných

informací z velmi rozsáhlých databází. Důležitou vlastností dolování dat je, že se jedná o analýzy odvozené z obsahu dat, nikoli analýzy předem specifikované uživatelem, a jedná se především o odvozování prediktivních informací nikoli pouze deskriptivních. Dolování dat slouží manažerům k objevování nových skutečností, čímž pomáhají zaměřit jejich pozornost na podstatné faktory podnikání, umožňují testovat hypotézy, odhalují ve stále se zrychlujícím a složitějším obchodním prostředí skryté korelace mezi ekonomickými proměnnými apod. (Gála, 2009, str. 230)

Dle L. Gály není data mining proces jen samotná analýza dat, ale jedná se především i o předpovídání a vyhledávání nových informací a poznatků z již existujících dat. Proces data mining tak můžeme vyložit jako podpůrný nástroj pro business intelligence, oba pojmy spolu bezesporu blízce souvisí a v některých případech mohou označovat to samé.

Příklady očekávaných výsledků použití data mining podle H. Skalské (2010):

- Model pro řízení vztahů se zákazníky (indikátory bonity zákazníka).
- Předvídání reakce zákazníka na plánované změny dodavatelů.
- Analýza účinnosti webových stránek na základě chování uživatelů webu (uspořádání stránek s cílem zlepšení finančních ukazatelů internetového obchodu).
- Předvídání rizika (včasná předpověď ztráty bonitních zákazníků na základě změny jejich obvyklého chování, návrh intervence).
- Klasifikace žadatelů o úvěr v bance podle jejich očekávané schopnosti splácet půjčku.
- Předpověď kladné odezvy na marketingovou nabídku na základě speciálních a demografických znaků oslovených osob.
- Identifikace podvodného jednání analyzováním nepříjemných indikátorů nebo výrazných odchylek od typických vzorů.
- Odhalování vztahů mezi nákupními zvyklostmi zákazníků za účelem cílené nabídky.

Ze seznamu je patrné široké množství využití data mining (business intelligence) nástrojů. Jejich zavedení ve společnosti může znamenat značnou konkurenční výhodu.

3.1.2 Metody data miningu

Podle E. Turbana (2011) existuje více druhů data miningových metod, z nichž každá je vhodná pro jiný typ úloh. Metody se dělí na **klasifikační, regresní, shlukové a asociační**.

3.1.2.1 Klasifikační metody

Klasifikační metody jsou asi nejpoužívanější metodou z výše zmiňovaných. Jedná se o metodu, která využívá tzv. *machine-learning*, tedy **učení z předešlých dat a událostí k odhadnutí stavů a chování v budoucnosti**. Klasifikační metody mohou být využity v širokém množství oborů. Příkladem jejich aplikování může být meteorologie a předpověď počasí (slunečné, deštivé, zamračené). V marketingu se tato metoda dá použít za účelem přímého cílení na zákazníka (získáme zákazníka či nikoliv) nebo například ve finančním sektoru pro odhadnutí budoucí ziskovosti klienta (ziskový, neziskový). Pokud potřebujeme **předpovědět konkrétní číselnou hodnotu** (teplota počasí bude 20°C), je nutné použít metodu **regresní**. (Turban, 2011, str. 178)

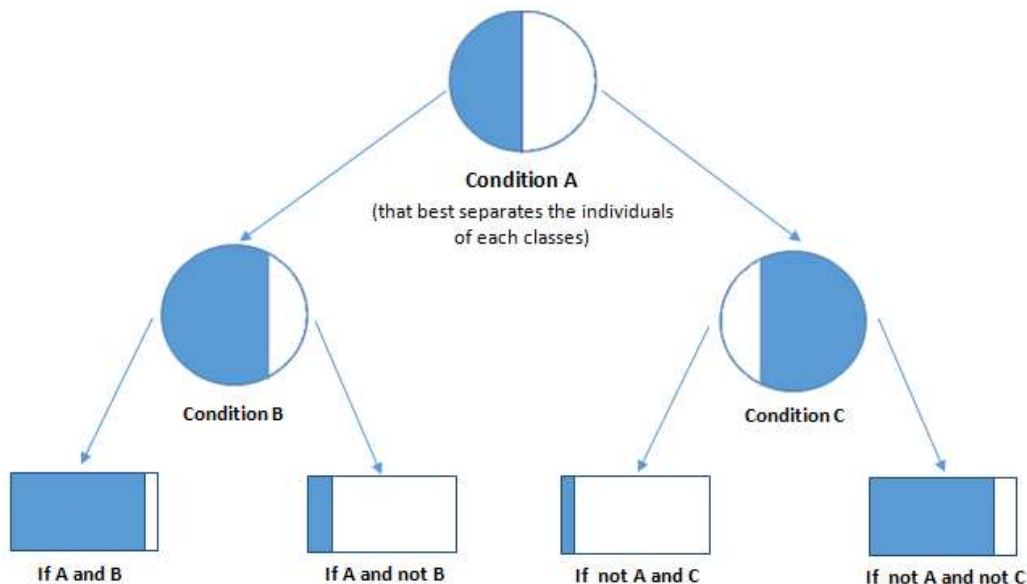
Rozhodovací stromy

Rozhodovací stromy jsou další populární klasifikační metodou. Základní myšlenkou je neustálé **rozdělování dat do menších uzlů** (node) **na základě podmínek**, které jsou schopny nejlépe daná data separovat. Tento proces pokračuje do té doby, dokud nějaký z uzlů neobsahuje pouze data stejné třídy nebo relativně malý počet dat, která už by už neměla smysl dále rozdělovat.

Celý proces i samotná data je vhodné rozdělit do tří částí: **trénovací část**, na které jsou data rozdělena do uzlů, dokud každý z nich nebude obsahovat jen data stejné třídy. **Testovací část**, která má za účel zobecnit celý model a **validační část** dat, která ověřuje a zlepšuje přesnost modelu.

Setkat se můžeme s různými algoritmy rozhodovacích stromů, nejznámějšími jsou C4.5, C5 a CHAID. (Turban, 2011, str. 182)

Obrázek č. 2: Rozhodovací strom



Zdroj: Upraveno dle Tuffery, 2011, str. 314

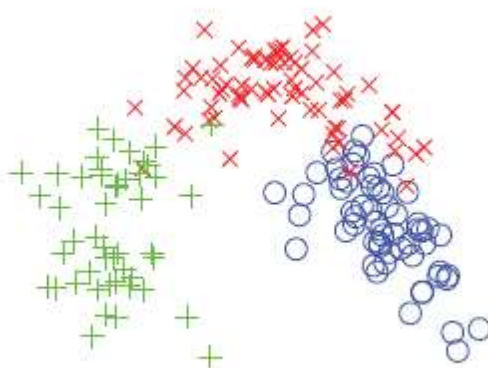
3.1.2.2 Shlukové metody

Shlukové metody se používají především k **nalezení podobných (homogenních) objektů**, které mohou být potom dále zkoumány různými způsoby. Této techniky se využívá v širokém množství oborů, jako například v biologii, lékařství, genetice, analýze sociálních sítí, antropologii, archeologii, astronomii a v neposlední řadě i v managementu.

Principem shlukových metod je **vytvoření shluků** (clusterů), kde objekty (lidé, věci, události), které do nich patří, mají **silnou vnitřní vazbu** (jsou si podobné) a naopak **externí vazba** (rozdíl mezi objekty z jiných shluků) **je co největší**.

Příkladem využití shlukových metod v praxi může být například rozdělení žáků do skupin s ohledem na jejich prospěch a chování, nebo rozdělení zákazníků do skupin pro vhodné zvolení marketingového přístupu.

Obrázek č. 3: Shlukové metody



Zdroj: Pividori, 2016

Shlukové analýzy mohou být například použity v jedné z následujících statistických metod: k-means, neuronové sítě, fuzzy-logika a další. (Turban, 2011, str. 184) (Tuffery, 2011, str. 235)

3.1.2.3 Asociační metody

Asociační metody jsou nejčastěji využívány ke zkoumání tzv. spotřebního koše. Spotřební koš vyjadřuje zboží, které si zákazník koupí při jedné návštěvě obchodu. Díky tomu se občas místo názvu asociační metoda používá název: **market basket analysis** (analýza spotřebního koše).

V literatuře se uvádí **příklad využití analýzy spotřebního koše** u jednoho nejmenovaného obchodního řetězce, který objevil závislost mezi zákazníky kupující pivo a zároveň dětské pleny. Tento jev bývá vysvětlován tím, že otcové (především mladí) používají nákup plen jako záminku pro návštěvu obchodního domu, aby mohli současně nakoupit pivo. Tímto zjištěním mohl obchodní řetězec zvýšit prodeje tím, že je umístil vedle sebe, nebo naopak od sebe, aby kupující musel projít celý obchodní dům. (Turban, 2011, str. 186)

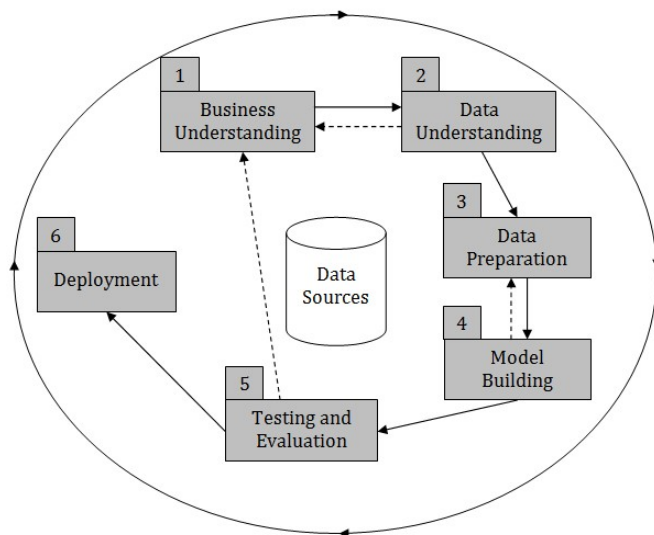
3.1.3 Proces data miningu

Z důvodu náročnosti celého data miningu procesu, byly zavedeny standardy, které celý proces rozdělují do několika bodů a zvyšují tak pravděpodobnost zdárného dosažení výsledků.

3.1.3.1 CRISP-DM

Jedním ze standardů je i **CRISP-DM** (Cross-Industry Standard Process for Data Mining), který byl poprvé navržen v roce 1996 Evropským konsorciem firem, jako neziskový standard pro data mining. Diagram je zobrazen na obrázku č. 4.

Obrázek č. 4: CRISP-DM proces



Zdroj: Upraveno dle Turban, 2011, str. 169

Standard CRISP-DM rozděluje data mining proces do šesti dílčích částí, které jsou blíže popsány níže. (Turban, 2011, str. 169)

1. Porozumění obchodní stránce

Prvním důležitým bodem je uvědomění si, k čemu má celý data mining projekt sloužit a to je především pro **potřebu společnosti**. Na začátku každého projektu by tak měl být jednoduchý cíl vycházející z managementu firmy. (př. Jaké jsou společné znaky zákazníků, kteří od nás odcházejí?). V této souvislosti je také nutné zvážit rozpočet na celý projekt. Nežádoucí je, aby náklady na zpracování převýšily potencionální hodnotu výsledků.

2. Porozumění datům

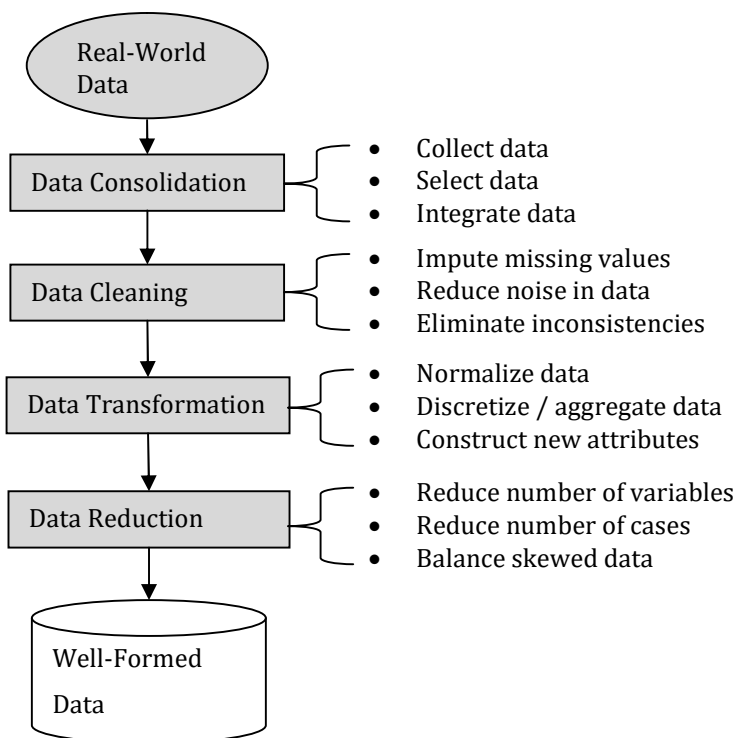
Pochopení dat je velice důležitý bod. Jen těžko lze zkoumat data, o kterých nevíme, co vyjadřují. Důležité tedy je, mít **reprezentativní data** (která souvisí s cílem data miningu), vědět **odkud pocházejí a jak byla sbírána**. Nesprávné zís-

kávání nebo pochopení dat může vést ke špatné interpretaci a ke znehodnocení celého projektu.

3. Příprava dat

Příprava dat je mezikrok mezi pochopením dat a modelováním. Pokud získáme data a chápeme jejich význam, můžeme je dále **upravovat a připravit na samotné modelování**. Tento krok zabírá až 80% času z celého projektu. Důvodem je především to, že data z reálného světa nejsou nikdy ideální. Postup transformace do dat, která mohou být dále využita k data miningu, je zobrazen na grafu č.5.

Obrázek č. 5: Kroky přípravy dat pro DM



Zdroj: Upraveno dle Turban, 2011, str. 171

Dělí se do čtyř kategorií: **sběr, čištění, transformace a redukce**. Každá z nich obsahuje ještě další složky. Můžeme zde tak najít získávání a selekci dat, doplnění chybějících hodnot, eliminaci nekonzistentních nebo extrémních hodnot, normalizaci, vytváření nebo odstraňování proměnných a další.

4. Modelování

Vzhledem k velikému množství různých modelovacích technik neexistuje universální postup při vytváření modelů. Jednotlivé modely navíc vyžadují různé parametry pro dosažení optimálních výsledků.

5. Testování a hodnocení

Modely jsou dále testovány a hodnoceny na základě jejich přesnosti a obecnosti. Nedílnou součástí je také kontrola, zda model naplňuje cíle, které byly do celého projektu vloženy. V případě nutnosti je nezbytné model doplnit, přepracovat nebo se vrátit do kroku č. 1 a začít od začátku s nově definovaným cílem projektu.

Další možností je otestovat správnost a funkčnost modelu „v reálném světě“. To ovšem záleží na konkrétním případě a možnostech rozpočtu.

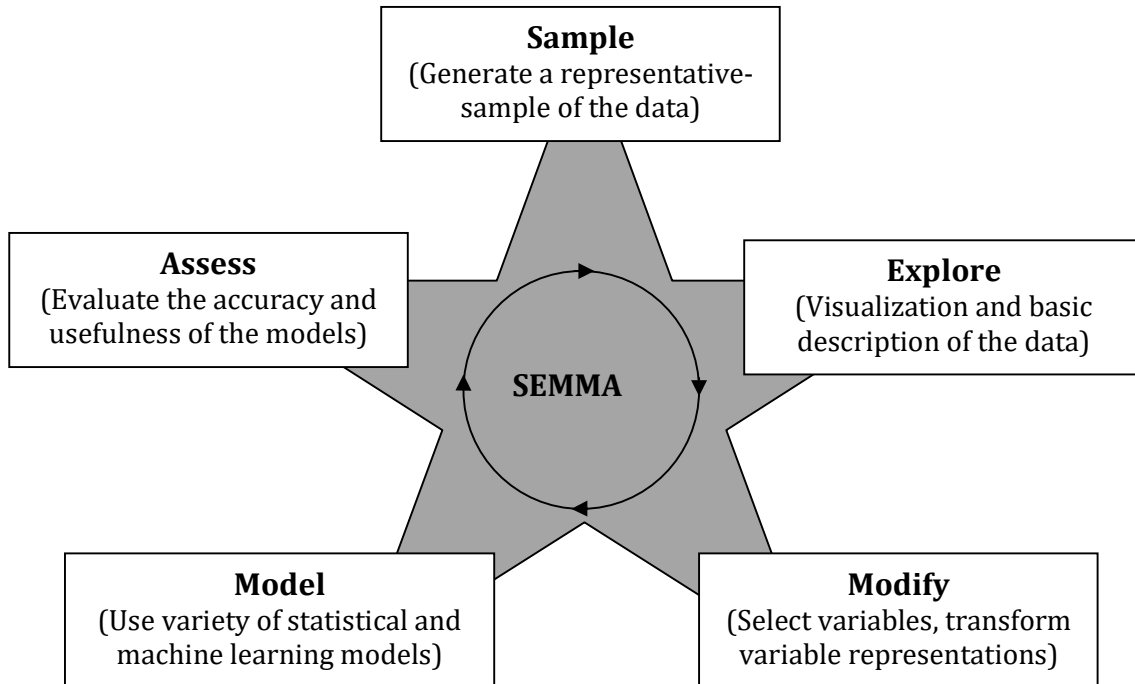
6. Použití

Poslední částí je získané poznatky správně interpretovat a vysvětlit, aby mohly být dále využity ku prospěchu. Může se jednat o jednoduchý report, ale i komplexní návrh řešení pro implementaci skrz celou společnost.

3.1.3.2 SEMMA

Dalším známým standardem je SEMMA, který byl vyvinut institutem SAS. Jedná se o akronym slov: „**sample, explore, modify, model, and assess**“. Hlavním rozdílem mezi CRISP-DM a SEMMA je ten, že CRISP-DM vkládá větší důraz na obsáhlejší pochopení celého procesu, především pochopení podnikových požadavků. Naopak SEMMA předpokládá, že cíle data mining procesu byly spolu se zdroji dat pochopeny a nevěnuje jim takovou pozornost. (Turban, 2011, str. 176)

Obrázek č. 6: SEMMA Data Mining proces



Zdroj: Upraveno dle Turban, 2011, str. 176

3.2 Finanční instituce

Finanční instituce se začaly vytvářet a formovat spolu s rostoucím významem obchodování ve společnosti a se zaváděním platidel. Z historického pohledu sloužily k placení nejdříve cenné kovy, jako například zlato a stříbro. Postupem času však tyto suroviny nestačily pokrýt potřeby obchodu, a tak se začaly využívat směnky. Už ve 14. století v Benátkách vznikli ochránci depozit a vypořádání dluhů se začalo provádět účetním převodem. O tři staletí později se Velká Británie díky založení Britské centrální banky stala centrem evropského obchodu a vhodným místem pro bezpečné investování kapitálu. Podpora obchodování spočívala především v akceptaci směnek a poskytování dlouhodobých prostředků pro průmysl. Začalo se také rozvíjet obchodování s bavlnou a vznikala speciální odvětví zabývající se financováním a pojištěním lodní dopravy. Pojišťovny rozšiřovaly svoje portfolio a získávaly velké množství finančních prostředků, které bylo nutno dále investovat. Začaly vznikat také stavební společnosti, které umožnily bezpečné uložení úspor a nabízely prostředky pro nákup domů.

Pod pojmem finanční instituce si tedy můžeme představit všechny subjekty, které obchodují nebo zprostředkovávají obchod na finančních trzích.

Finanční zprostředkování má mnoho společného se zprostředkováním reálných aktiv. Nakupovat a prodávat na finančních trzích je možné v podstatě stejnými způsoby jako na trhu reálných aktiv, tj. nemovitostí, automobilů apod. Investiční aktivita je výsledkem procesu, kdy finanční instituce jako zprostředkovatelé ve finančním sektoru řídí svá portfolia úvěrů, půjček a cenných papírů. Obchodní a investiční banky, stavební spořitelny, pojišťovny, investiční společnosti, investiční fondy a penzijní fondy provádějí finanční transakce na svůj i cizí účet. Obchodní banky berou od vkladatelů depozita. Investiční banky fungují na primárním trhu s cennými papíry. Pojišťovací trh zajišťuje určitou úroveň bezpečnosti pro reálný sektor. Stavební spořitelny umožňují klientům výhodné úvěry na financování bytových potřeb a také výhodné investiční možnosti. Penzijní fondy poskytují dnes malým investorům možnost ukládat prostředky za účelem budoucí spotřeby. Pojišťovny a penzijní fondy investují své prostředky ve finančním sektoru. Většinu investic na kapitálových trzích vytváří finanční instituce. (Jílek, 1997, str. 423)

3.2.1 Funkce finančních institucí na finančním trhu

V této kapitole jsou popsány druhy finančních institucí vyskytujících se na území ČR. Zároveň je popsán důvod jejich vzniku a jejich místo na finančním trhu.

Podle S. Poloučka (2009) můžeme rozdělit finanční instituce podle několika kritérií.

Základním kritériem je dělení podle toho, jestli poskytují svým klientům depozitní služby, či nikoliv. Depozitní službou se myslí možnost uložení finančních prostředků ve formě depozit na požádání, NOW (negotiable order of withdrawal) účtů a účtu úspor.

Mezi **depozitní finanční instituce** řadíme:

- komerční banky,
- spořitelny,
- vzájemné spořitelny a úvěrová družstva.

Mezi **nedepozitní finanční instituce** řadíme:

- peněžní tržní fondy,
- pojišťovny,
- penzijní fondy,
- finanční společnosti,
- investiční společnosti,
- obchodníky s cennými papíry.

Nesmíme ale opomenout centrální banku, která operuje nad finančním trhem a zajišťuje dohled nad všemi finančními institucemi v dané zemi.

V posledních letech se stále častěji setkáváme s prolínáním jednotlivých druhů finančních institucí, včetně produktů, které nabízejí. Běžně se tak můžeme setkat s depozitními finančními institucemi, které nabízejí nedepozitní produkty, jako doplněk svých běžně poskytovaných služeb.

3.2.1.1 Centrální banky

Centrální banky jak je známe nyní, začaly vznikat v polovině 19. století, v řadě zemí se však založení datuje až po 2. světové válce. V samostatné České republice tuto funkci zastává Česká národní banka (ČNB) od roku 1993.

Základními funkcemi centrální banky v tržní ekonomice jsou:

- emisní funkce,
- funkce vrcholného subjektu měnové politiky,
- funkce regulace bankovního systému,
- a také zajišťování tzv. *banky bank a státu*.

Emisní funkcí je myšleno emitování hotovostních peněz (bankovky, mince) na daném území. Emisním právem se centrální banka stává monopolem a má tak schopnost určit a kontrolovat množství peněz v oběhu. Se vznikem bezhotovostních peněz se však možnosti centrální banky v této oblasti omezují.

Centrální banka je **vrcholným subjektem měnové politiky**, která má za cíl dosahování stanovených cílů, především cenovou stabilitu dané měny a podporu hospodářské politiky vedoucí k udržitelnému hospodářskému růstu. Rozumí se tím především nastavení úrokových sazeb, které budou držet inflaci na nízké a stabilní úrovni, aniž by se tím zpomalilo tempo růstu ekonomiky nebo naopak přehnaně urychlovalo.

Další významnou úlohou je **regulace bankovního systému**. Centrální banka prosazuje pravidla pro činnosti bankovních institucí v dané zemi a kontroluje jejich dodržování. Rozhoduje tak o přidělení nebo sebrání bankovní licence.

Centrální banka vede účty ostatním bankám, poskytuje jim úvěry, spravuje jejich vklady a provádí platební a zúčtovací operace. V literatuře se setkáme s pojmem **banka bank**.

Její další funkcí je vykonávání **banky státu**, kdy spravuje účet vlády a zajišťuje vypořádání pro běžné hospodaření státního rozpočtu.

Centrální banka vykonává i další funkce, jako například správu devizových rezerv, ochranu a informování bankovních i nebankovních subjektů o svých přístupech a postupech a v neposlední řadě vystupuje i jako reprezentant dané země

ve vztahu k zahraničí. (Revenda, 2012, str. 213-218), (Česká národní banka. *O ČNB – Česká národní banka. [online]. 2016 [cit. 2016-01-19]. Dostupné z: https://www.cnb.cz/cs/o_cnb/*)

3.2.1.2 Obchodní banky

Obchodní banky jsou jedním z nejvýznamnějších finančních zprostředkovatelů. Jejich právní vymezení lze nalézt v *zákoně o bankách*, který vychází z předpisů Evropské unie. Zákon vymezuje čtyři základní podmínky, které obchodní banka musí splňovat:

- jedná se o **akciovou společnost** se sídlem v České republice;
- přijímají **vklady** od veřejnosti (které představují závazek vůči vkladateli na jejich výplatu);
- poskytuje **úvěry**;
- má udělenou **bankovní licenci**, kterou uděluje Česká národní banka.

Banka může vykonávat i další činnosti, má-li je povoleny v bankovní licenci. Jsou to například tyto: investování do cenných papírů na vlastní účet, finanční leasing, platební styk a zúčtování, vydávání platebních karet a šeků, poskytování záruk, finanční makléřství, směnářská činnost, pronájem bezpečnostních schránek. (Zákon č. 21/1992 Sb, o bankách)

Mezi základní funkce obchodní banky můžeme označit:

- finanční zprostředkování,
- emise bezhotovostních peněz,
- provádění platebního styku,
- zprostředkování finančního investování na peněžním a kapitálovém trhu.

Obchodní banka zajišťuje **přerozdělení finančních prostředků** od přebytkových k deficitním jednotkám. Podnikatelské subjekty tak mohou přemístit získané finanční prostředky tam, kde přinášejí nejvyšší zhodnocení.

Centrální banka je jediná, která může emitovat hotovostní peníze v daném státě. Bezhotovostní peníze však **mohou emitovat** i obchodní banky a to nad rámec přijatých depozit.

Provádění platebního styku znamená převod bezhotovostních peněz z účtu svých klientů na jiný účet bez nutnosti hotovostních peněz. Obchodní banky také mohou pro své klienty zajišťovat nákup cenných papírů, případně jiné **investiční obchody**. (Revenda, 2012, str. 87)

Bankovní produkty podle Z. Revendy (2012) můžeme rozdělit do následujících kategorií.

Bankovní finančně úvěrové produkty, které umožňují získat finanční prostředky od banky:

- kontokorentní úvěr,
- eskontní úvěr,
- hypoteční úvěr,
- spotřebitelský úvěr,
- bankovní záruky,
- faktoring,
- forfaiting.

Depozitní (vkladové) bankovní produkty, které slouží pro finanční investování:

- vklady na viděnou,
- termínované vklady,
- úsporné vklady,
- bankovní dluhové cenné papíry,
- stavební spoření.

Platebně zúčtovací bankovní produkty:

- příkazy k úhradě a inkasu (bankovní převody),
- šeky,
- platební karty,

- dokumentární akreditiv,
- dokumentární inkaso.

3.2.1.3 Spořitelny

*Důvody pro vznik spořitelen (savings and loan associations) byly velice prozaické – až do zhruba 50. a 60. let bylo obtížné získat u banky dlouhodobý spotřebitelský úvěr, např. úvěr na koupi domu. Banky poskytovaly především krátkodobé úvěry a měly zájem o podnikatelskou klientelu. První spořitelny vznikaly jako vzájemné spořitelny (manual savings banks) na počátku 19. století v USA a ve Skotsku (zde jsou nazývány trustee saving banks). Byly založeny na myšlence **shromáždit úspory drobných střadatelů**, spojit je a uložit v bance. Úrokový výnos pak byl rozdělen mezi podílníky, proto také kdo uložil prostředky, je považován za spoluvlastníka a takto uložená depozita jsou nazývána podíly (shares). Z hlediska vlastního fungování není podstatný rozdíl mezi spořitelny a vzájemnými spořitelny až na to, že vzájemné spořitelny fungují většinou jako družstva, zatímco spořitelny jako akciové společnosti. (Polouček, 2009, str. 304)*

Jedním z důvodů, proč se spořitelny staly populární a našly si svoje místo na finančním trhu, bylo nabízení **stavebního spoření**. Jedná se o účelový druh spoření, při kterém vkladatel dlouhodobě ukládá prostředky. Celé spoření je navíc doplněno státní podporou a úrokem. Po uplynutí doby spoření má vkladatel nárok na výhodný úvěr ze stavebního spoření na bytové potřeby.

3.2.1.4 Pojišťovny a penzijní fondy

Pojišťovny a penzijní fondy můžeme označit jako smluvní finanční společnosti, protože získávají své finanční prostředky uzavřením kontraktu s klienty. Jedná se většinou o dlouhodobé smlouvy, které dávají pojišťovnám a penzijním fondům **stabilní a relativně jisté finanční příjmy**. Tento fakt se také odráží ve způsobu jejich investování, ve kterém se zaměřují především na **výnosnější, dlouhodobé cenné papíry**. V malé míře také udržují část aktiv v likvidní formě, pro nečekané události a výdaje.

Smluvní finanční instituce můžeme rozdělit do dvou základních skupin. První skupinou jsou **pojišťovny životní** (krytí rizika úmrtí a dožití) a **pojišťovny**

neživotní (pojištění majetku, živelné pojištění a pojištění odpovědnosti za škodu). Druhou skupinou jsou **penzijní fondy**, které shromažďují finanční prostředky svých klientů, které následně investují a po vzniku nároku vyplácejí zpět. (Polouček, 2009, str. 309)

Pojišťovny mají za cíl krýt rizika spojená s běžnými činnostmi člověka, od zranění až po majetkové, živelné pohromy a smrt. Pomáhají tak zajistit jedince a případně i společnost před nečekanými událostmi, které mohou nastat.

Penzijní fondy slouží jako prostředek spoření, pro zajištění stávající úrovně života v důchodovém věku.

3.2.1.5 Investiční bankovníctví

Podstata investičního bankovníctví spočívá v **přesunu úspor** (přebytkových jednotek) do jednotek deficitních. V praxi to pak například spočívá ve vydávání cenných papírů na straně deficitní a v jejich následném nákupu stranou přebytkovou. Tím dochází k alokaci finančních zdrojů, kde přinášejí vyšší zhodnocení.

Podobnou funkci ale vykonávají i komerční banky, rozdíl mezi investičním bankovníctvím a komerčními bankami popisuje Revenda takto:

*Investiční bankovníctví obsahuje především **zprostředkovatelské obchody na účet svých zákazníků**. V komerčním bankovníctví (commercial banking) naopak převládají obchody na vlastní účet a riziko, přičemž hlavní zdroj zisku tvoří úroková marže mezi aktivními a pasivními úrokovými obchody. V oblasti investičního bankovníctví je **hlavním pramenem zisku zprostředkovatelská provize, popř. rozdíl mezi nákupním a prodejním kursem**.*(Revenda, 2012, str. 115)

P. Musílek (2002) definoval charakteristické rozdíly mezi investičními zprostředkovateli a komerčními bankami viz. tabulka č.1.

U investičních společností je charakteristické, že výnos nezávisí na velikosti jejich aktiv, jako v případě bank komerčních, ale především na objemu transakcí a množství zprostředkovaných obchodů.

Tabulka č. 1: Definiční charakteristiky investičních zprostředkovatelů a komerčních bank

Charakteristika	Investiční zprostředkovatel	Komerční banka
Instrumenty	cenné papír, finanční deriváty	depozita, úvěry
Druh obchodů	mimobilanční	bilanční
Zdroj zisku	provize, kapitálové zisky	úroková marže
Riziko	nedostatek klientů, tržní riziko	úvěrové riziko, úrokové riziko
Financování	REPO	primární depozita

Zdroj: Musílek, 2002, str. 127

Z. Revenda (2012) dále vymezil investiční bankovníctví jako obchody s cennými papíry, finančními deriváty a ostatními instrumenty trhu cenných papírů. Do investičního bankovníctví zahrnul:

- emisní obchody,
- vlastní a zprostředkovatelské obchody s investičními instrumenty,
- depotní obchody,
- majetkovou správu,
- fúze a akvizice.

P. Musílek (2002) do svého vymezení zahrnul ještě další aktivity, týkající se vydávání, obchodování, analyzování a správy investičních instrumentů. Zejména emisní obchody, obchody na sekundárních trzích, investiční poradenství, investiční analýzy, správu aktiv, finanční inženýrství a zprostředkování fúzí a akvizic.

3.2.2 Začlenění společnosti Akcenta CZ, a.s. na finanční trh

Dle klasifikace institucionálních sektorů a subsektorů (CZ-CISS) dle ESA (Evropský metodologický standard) 2010 se společnost Akcenta CZ, a.s. řadí mezi: **Ostatní finanční zprostředkovatele kromě pojišťovacích společností a penzijních fondů, národní soukromí.** (Registr ekonomických subjektů, online)

Z hlediska předmětu podnikání poskytuje společnost platební služby dle zák. č. 284/09sb.:

- **vkládání hotovosti na platební účet** vedený společností dle § 3 odst. 1 písm. a) uvedeného zákona,
- **výběr hotovosti z platebního účtu** vedeného společností dle § 3 odst. 1 písm. b) uvedeného zákona,
- provádění **převodu peněžních prostředků** dle § 3 odst. 1 písm. c) uvedeného zákona,
- **vydávání a správa platebních prostředků** a zařízení k přijímání platebních prostředků dle § 3 odst. 1 písm. e) uvedeného zákona,
- **provádění převodů peněžních prostředků** dle § 3 odst. 1 písm. f) uvedeného zákona

a činnost obchodníka s cennými papíry dle zák. č. 256/04 Sb. v rozsahu hlavních investičních služeb, tj.:

- **přijímání a předávání pokynů týkajících se investičních nástrojů,**
- **provádění pokynů týkajících se investičních nástrojů** na účet zákazníka,
- **obchodování s investičními nástroji** na vlastní účet,
- **doplňkové investiční služby:** úschova a správa investičních nástrojů včetně souvisejících služeb. Vše ve vztahu k investičním nástrojům podle § 3 odst. 1 písm. d) uvedeného zákona (Veřejný rejstřík a Sbírka listin, online)

S pohledu zákona, tak může společnost Akcenta CZ, a.s. vykonávat platební služby, podobně jako **komerční banky**, tedy především operace s platebním účtem klienta. Zároveň je však schopna provádět operace týkající se investičních nástrojů, jako v případě **investičních bank**. Jedná se tedy o kombinaci více typů finančních institucí.

V praxi tak Akcenta CZ, a.s. poskytuje služby týkající se **zahraničních plateb, směnu deviz a zajištění kurzového rizika**. K zajištění kurzového rizika dochází za pomoci termínovaných obchodů. Na výběr má klient pět alternativ:

- forwardy,
- forwardy s klouzavou dobou splatnosti,
- amortizační forward,

- devizový swap,
- měnová FX opce.

(Zahraniční měny a platby – AKCENTA CZ, online)

Začlenění společnosti Akcenta CZ, a.s. mezi typové finanční instituce není možné, jedná se totiž o případ propojení více typů finančních institucí, se zaměřením na mezinárodní obchod a devizové operace, i proto ji Český statistický úřad řadí mezi ostatní finanční zprostředkovatelé.

Společnost Akcenta CZ, a.s. je ukázkou moderního trendu, kdy dochází k prolínání služeb různých typů finančních institucí a k specializaci na určitý druh poskytování služeb.

4 Aplikace BI nástrojů ve finanční instituci

Od zástupce finanční firmy AKCENTA CZ, a.s. byl získán datový soubor, který obsahoval informace o klientech a jejich uskutečněných transakcích. Odpadla tak starost se získáváním dat z jiných informačních systémů a jejich následném setřídění do souvislého datového souboru.

Vzhledem k povaze a náročnosti vytyčených cílů bude pro analýzu použit **statistický nástroj od firmy IBM - SPSS Modeler**. Jedná se o statistický software, sloužící k predikování, analyzování a podpoře rozhodování. Umožňuje provádění komplexních analytických operací bez nutnosti jejich hlubších znalostí a snadnou manipulaci s daty. Na jednodušší operace bude použit **Microsoft Excel**, který vyniká svým zažitým ovládáním a jednoduchým zobrazením dat do grafické podoby – grafů.

Hlavním cílem praktické části je odhalit bonitního klienta na začátku smluvního vztahu. Nejdříve je nutné současné klienty zařadit do skupin podle jejich sklonu k budoucí ziskovosti pro společnost. Teprve poté můžeme zkoumat společné znaky a sestavit model. Součástí zkoumání je také nezbytná příprava dat, případně vytváření nových proměnných z těch stávajících.

Postup práce, je pro přehlednost rozdělen do více částí:

1. popis datového souboru,
2. rozdělení stávajících klientů do skupin podle jejich ziskovosti,
3. příprava dat (vytvoření nových proměnných),
4. analýza klíčových proměnných,
5. odhalení bonitního klienta na začátku smluvního vztahu (sestavení modelu).

4.1 Datový soubor

Zdrojem dat je excelový soubor .xlsx, čítající záznamy o 2 594 klientech. Data jsou rozdělena do tří sad:

- evidenční a sumární údaje,
- data o prvních pěti transakcích klientů,

- data o prvních pěti transakcích klientů omezená na dobu prvních třech měsíců od podpisu smlouvy.

Důvodem třetí datové sady a odlišení na dobu prvních třech měsíců od podpisu smlouvy je ten, že zákazníci mají zpočátku zvýhodněný sazebník, který může ovlivnit výsledné chování klienta.

Každá datová sada obsahuje několik proměnných, rozdělených podle doby uskutečnění. Mezi základní sledované atributy patří objem klientských transakcí, profit za určité období nebo za jednotlivé transakce, počet uskutečněných transakcí a doba mezi jejich uskutečněním. Data jsou dále rozdělená na krátkodobá – za první tři měsíce od uzavření smlouvy, střednědobá – za první půl rok a dlouhodobá za první rok od uzavření smlouvy.

1. Datová sada - evidenční a sumární údaje

- **Klient-ID** – evidenční údaj
- **Klient-právní forma** – údaj o právní formě klienta (0-právnícká osoba, 1-fyzická osoba podnikatel, 2-fyzická osoba nepodnikatel)
- **Objem_1-3, Objem_1-6, Objem_7-12, Objem_1-12** – objem klientských transakcí od uzavření smlouvy (1. čtvrtletí, 1. pololetí, 2. pololetí, celý rok)
- **Profit_1-3, Profit_1-6, Profit_7-12, Profit_1-12** – profit z klientských transakcí (1. čtvrtletí, 1. pololetí, 2. pololetí, celý rok)
- **Počet_tr_1-3, Počet_tr 1-6, Počet_tr 7-12, Počet_tr 1-12** – počet klientských transakcí (1. čtvrtletí, 1. pololetí, 2. pololetí, celý rok)
- **Profit_na_tr_1-3, Profit_na_tr_1-6, Profit_na_tr_7-12, Profit_na_tr_1-12** – průměrný profit na transakci z klientských transakcí (1. čtvrtletí, 1. pololetí, 2. pololetí, celý rok)

2. Datová sada – data o prvních pěti transakcích klientů

- **RS - 1_ob**-doba mezi podpisem smlouvy a prvním obchodem ve dnech
1 - 2_obch- doba mezi prvním a druhým obchodem ve dnech
2 - 3_obch- doba mezi druhým a třetím obchodem ve dnech
3 - 4_obch- doba mezi třetím a čtvrtým obchodem ve dnech

- 4 - **5_obch**- doba mezi čtvrtým a pátým obchodem ve dnech
 - 5 - **6_obch**- doba mezi pátým a šestým obchodem ve dnech
 - **1_profit, 2_profit, 3_profit, 4_profit, 5_profit**- profit z jednotlivých transakcí (1. až 5. transakce)
 - **1_objem, 2_objem, 3_objem, 4_objem, 5_objem**- objem z jednotlivých transakcí (1. až 5. transakce)
- 3. Datová sada - Data o prvních pěti transakcích klientů omezená na dobu prvních třech měsíců od podpisu smlouvy**
- **1_profit_3, 2_profit_3, 3_profit_3, 4_profit_3, 5_profit_3** - profit z jednotlivých transakcí omezený na dobu 3 měsíců od podpisu smlouvy (1. až 5. transakce)
 - **1_objem_3, 2_objem_3, 3_objem_3, 4_objem_3, 5_objem_3** - objem z prvních pěti transakcí
 - **RS - 1 ob** - doba mezi podpisem smlouvy a prvním obchodem ve dnech
 - 1 - **2_obch_3** - doba mezi prvním a druhým obchodem ve dnech
 - 2 - **3_obch_3** - doba mezi druhým a třetím obchodem ve dnech
 - 3 - **4_obch_3** - doba mezi třetím a čtvrtým obchodem ve dnech
 - 4 - **5_obch_3** - doba mezi čtvrtým a pátým obchodem ve dnech

4.2 Rozdělení klientů

Klienty je nejprve nutné rozdělit do skupin podle jejich přínosů pro firmu. Ne všichni zákazníci přinášejí firmě zisk, a proto je pro společnost nevýhodné je dlouhodobě udržovat. Tato skupina zákazníků by měla být co nejdříve odhalena a buď jejich ztrátová aktivita vykompenzována vyššími poplatky a maržemi nebo od další spolupráce úplně odražena.

Zbylou (ziskovou) skupinu klientů můžeme dále rozdělit podle toho, jestli přinášejí dlouhodobý nebo pouze krátkodobý zisk.

Máme tedy určeny tři základní skupiny klientů:

- **ne profitabilní,**
- **krátkodobě profitabilní,**

- **dlouhodobě profitabilní.**

K rozdělení na profitabilní a ne profitabilní klienty je nutné znát, jaké náklady pro společnost přinášejí jejich jednotlivé transakce. Náklady ovlivňuje mnoho faktorů – kam zákazník posílá peníze, od jaké banky, jaký má sazebník a další. Tyto faktory se navíc v čase neustále mění a tak je jednoznačné stanovení nákladů prakticky nemožné. V průměru se jedná o 120 - 140 CZK na transakci. Po započítání veškerých nákladů firmy je hodnota cca. 400 CZK.

Klienti s dlouhodobým průměrem profitu na transakci pod 120 CZK jsou tedy proděleční a je nevhodné je dále udržovat.

Mezi 120 – 400 CZK se jedná o klienty, kde je nutné zvážit další faktory (například množství uskutečněných obchodů).

Nad 400 CZK se jedná o profitabilní klienty.

Na rozdělení klientů do jednotlivých skupin byl použit program Microsoft Excel, kde je za pomoci funkce „KDYŽ“ vytvořen rozhodovací mechanismus. Skupiny jsou označeny číselně a po všech úpravách budou vyjadřovat požadované skupiny klientů.

- 0 – ne profitabilní
- 1 – krátkodobě profitabilní
- 2 – dlouhodobě profitabilní

Rozhodovací mechanismus je pro větší přehlednost rozdělen do 3 částí:

- **První rozdělení klientů** (rozčlenění klientů na profitabilní, ne profitabilní a přechodnou skupinu, kde je nutné uvážit ještě počet uskutečněných transakcí).
- **Druhé rozdělení klientů** (rozdělení přechodné skupiny zákazníků podle počtu uskutečněných transakcí na profitabilní a ne profitabilní).
- **Konečné rozdělení klientů** (rozdělení profitabilních zákazníků na krátkodobě nebo dlouhodobě profitabilní).

4.2.1 První rozdělení klientů

Pro prvotní a hlavní rozdělení je použita proměnná „Profit na tr1-12“, která kombinuje celkový profit zákazníka za prvních 12 měsíců s počtem jeho uskutečněných transakcí (*Profit 1-12 / Počet tr 1-12*).

=KDYŽ(Profit na tr 1-12<120;0;KDYŽ(Profit na tr 1-12>400;2;1))

Když zákazníkuv profit na transakci za prvních 12 měsíců je menší než 120, je zařazen do skupiny 0 – ne profitabilní, když je větší, než 400 je zařazen do skupiny s označením 2 - profitabilní a pokud je mezi (120-400) je zařazen do skupiny 1, která je zatím určena jako přechodná pro další rozdělení.

Tabulka č. 2: První rozdělení (počty klientů)

	1. Rozdělení
0 - Ne profitabilní	153
1 - Přechodná skupina	228
2 - Profitabilní	2 207
Celkem	2 588

Zdroj: Vlastní zpracování

4.2.2 Druhé rozdělení klientů

U skupiny 1 – přechodná skupina je dále nutné uvážit počet transakcí klienta. Pokud se jedná o pravidelně obchodujícího klienta, je pro společnost výhodnější mu snížit transakční náklady, než o něj úplně přijít. Klient, který služby využívá minimálně, je pro společnost z dlouhodobého hlediska méně důležitý.

K určení hranice počtu transakcí je použita střední hodnota počtu transakcí za 12 měsíců.

Tabulka č. 3: Průměr a střední hodnota počtu transakcí

	Počet tr 1-3	Počet tr 1-6	Počet tr 7-12	Počet tr 1-12
Průměr	3,675	6,545	5,043	11,588
Střední hodnota	2	3	1	5

Zdroj: Vlastní zpracování

=KDYŽ(1.Rozdělení=1;KDYŽ(Počet tr1-12>=Střední hodnota(Počet tr 1-12);2;0);1.Rozdělení

Když první rozdělení (proměnná, kterou jsme získali předešlým krokem) je rovno 1 (klienti, kteří jsou na rozmezí a musíme u nich uvážit počet skutečných transakcí), a zároveň je počet transakcí za prvních 12 měsíců větší nebo roven střední hodnotě, jsou klienti přiřazeni do skupiny 2 (*profitabilní*). Pokud tomu tak není, řadí se do skupiny 0 (*ne profitabilní*). Jedná-li se o klienta, který není ve skupině 1, zůstává mu hodnota z prvního rozdělení klientů.

Tímto jsme rozčlenili klienty z přechodné skupiny 1 na *profitabilní*(2) nebo *ne profitabilní* (0).

Tabulka č. 4: Druhé rozdělení (počty klientů)

	1. Rozdělení	2. Rozdělení	Rozdíl
0 - Ne profitabilní	153	361	208
1 - Přechodná skupina	228	0	-228
2 - Profitabilní	2207	2227	20
Celkem	2588	2588	2588

Zdroj: Vlastní zpracování

Konkrétně se jednalo o **přeřazení 228 klientů na *ne profitabilní* a 20 na *profitabilní*.**

4.2.3 Konečné rozdělení klientů

Posledním postupem je rozčlenění profitabilních zákazníků na krátkodobě nebo dlouhodobě profitabilní. Pro klienty, kteří jsou profitabilní v krátkém i dlouhém období, budeme spíše volit označení *dlouhodobě profitabilní*, protože dlouhodobá výnosnost má pro společnost větší vypovídající hodnotu. Vzhledem k tomu, že do skupiny profitabilních zákazníků se dostali i zákazníci s nižším profitem na transakci než stanovených 400 Kč, provedeme porovnáním krátkodobého profitu na transakci (*Profit na tr 1-3*) s dlouhodobým profitem na transakci (*Profit na tr 1-12*).

Pokud se bude rovnat krátkodobý profit s dlouhodobým, přiřadíme klienta jako *krátkodobě profitabilního*, protože to bude ve většině případů znamenat, že od prvních 3 měsíců neuskutečnil žádné další transakce.

=KDYŽ(2.Rozdělení=2;KDYŽ(Profit na tr 1-3>=Profit na tr 1-12;1;2);2. Rozdělení)

Když druhé rozdělení je rovno 2 (pouze profitabilní zákazníci), jejich krátkodobý profit na transakci (za první 3 měsíce) je větší nebo roven dlouhodobému profitu na transakci (za prvních 12 měsíců) a je přiřazen do skupiny 1 (*krátkodobě profitabilní*). V jiném případě do skupiny 2 (*dlouhodobě profitabilní*). Pokud není splněna první podmínka, ponechá se hodnota druhého rozdělení (platí pouze pro skupinu 0 – *ne profitabilní*).

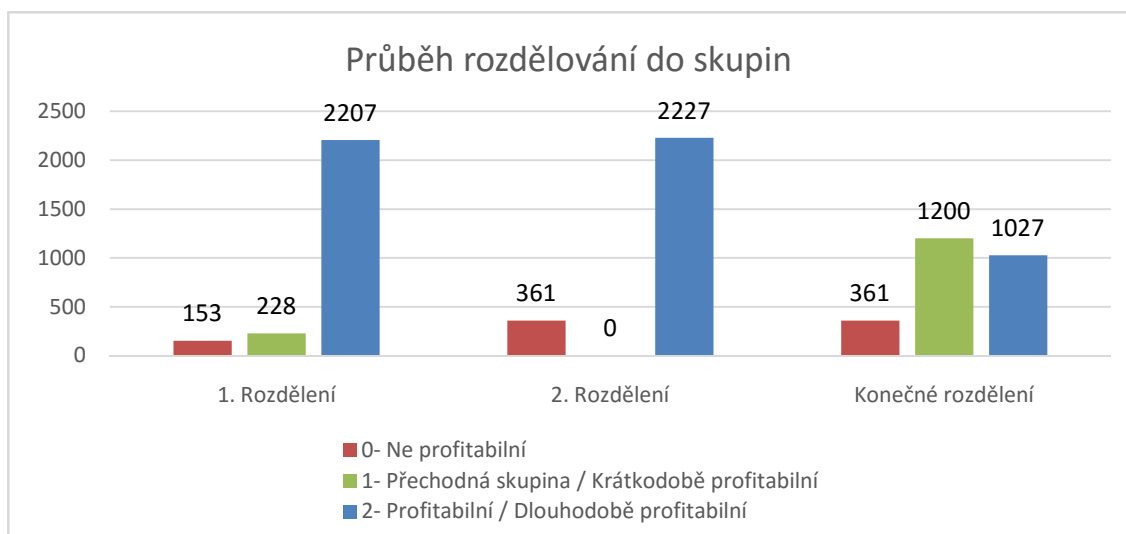
Tabulka č. 5: Konečné rozdělení (počty klientů)

	2. Rozdělení	3. Rozdělení	Rozdíl
0 - Ne profitabilní	361	361	0
1 - Krátkodobě profitabilní	0	1200	1200
2 - Dlouhodobě profitabilní	2227	1027	-1200
Celkem	2588	2588	2588

Zdroj: Vlastní zpracování

Třetí rozčlenění **přerozděluje profitabilní klienty na dlouhodobé a krátkodobé. Jako *krátkodobě profitabilní* jsme označili 1 200 klientů a jako *dlouhodobě profitabilní* 1 027.**

Graf č. 1: Počty klientů při přerozdělování do skupin



Zdroj: Vlastní zpracování

Graf označuje, jak se měnily počty klientů v průběhu rozdělování do požadovaných skupin. **Konečně rozdělení obsahuje 361 *ne profitabilních*, 1200 *krátkodobě profitabilních* a 1027 *dlouhodobě profitabilních* zákazníků.**

4.3 Vytvoření nových proměnných

Pro lepší manipulaci s daty a následné zkoumání je vhodné vytvořit si dodatečné proměnné z původních dat.

4.3.1 Označení jednorázových (one-time) klientů

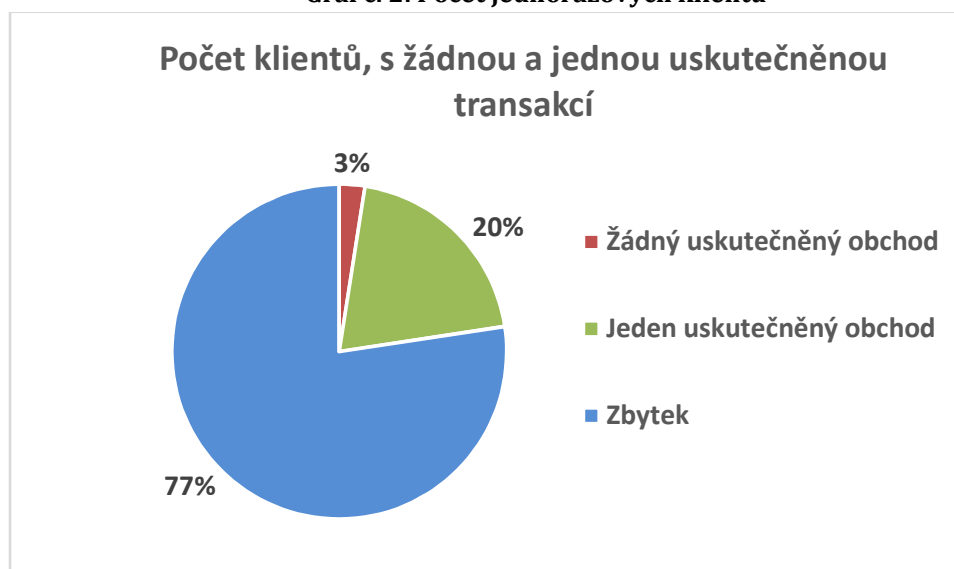
V datech se nachází velké množství klientů, kteří uskutečnili pouze jeden nebo žádný obchod. Na základě jedné nebo žádné uskutečněné transakce není vhodné usuzovat o profitabilitě, či ne profitabilitě klienta, a proto si tyto zákazníky označíme, abychom je poté mohli použít k dalšímu zkoumání.

=KDYŽ (Počet tr 1-12<=1;1;0)

Když je počet transakcí za prvních 12 měsíců menší nebo roven jedné, přiřadíme 1, v opačném případě přiřadíme 0.

Jedničkou tak máme označeny klienty, kteří uskutečnili jeden nebo žádný obchod za dobu 12-ti měsíců od podpisu smlouvy. **Celkově se jednalo o 521 klientů, kteří uskutečnili pouze jeden obchod a dalších 64, kteří neuskutečnili žádný.** Jedná se tedy o 585 zákazníků, které jsme označili v proměnné *one-time* jedničkou.

Graf č. 2: Počet jednorázových klientů



Zdroj: Vlastní zpracování

Z koláčového grafu lze lehce vyčíst, že se jedná skoro o čtvrtinu všech klientů (23%). **Společnost by se měla snažit odhalit, co je důvodem jejich neaktivity a snažit se klienty motivovat k uskutečnění dalších obchodů.**

4.3.2 Sumarizace objemu a profitu

Data v 3. datové sadě (data omezena na tři měsíce od podpisu smlouvy a na prvních 5 transakcích) jsou rozdělena podle jednotlivých transakcí (první až pátá). Vzhledem k tomu, že ne všichni klienti uskutečnili všech 5 transakcí, bude vhodné data sumarizovat než je zkoumat jako jednotlivé transakce. Tím se vyhneme i problému, jak by následně software naložil s chybějícími hodnotami, které přímo neznamenaají chybějící hodnotu, ale pouze neuskutečněnou transakci.

Byly vytvořeny nové proměnné **1_5_profit_3** a **1_5_objem_3** které jsou sumou pěti transakcí v horizontu tří měsíců od podpisu smlouvy.

```
1_5_objem_3 =SUMA(1_objem_3 : 5_objem_3)
```

```
1_5_profit_3 =SUMA(1_profit_3 : 5_profit_3)
```

4.3.3 Průměrná doba mezi transakcemi

Další nově vytvořenou proměnnou je **Prumer_transakci**, která označuje průměrnou dobu mezi prvními pěti uskutečněnými transakcemi.

```
Prumer_transakci=PRŮMĚR(RS-1obch :4-5 obch 3)
```

4.3.4 Nejmenší a největší transakce

Z prvních pěti transakcí klienta vyčleníme největší a nejmenší z nich.

Max_objem (největší klientova transakce)

```
MAX(1_objem_3 : 5_objem_3)
```

Min_objem (nejmenší klientova transakce)

```
MIN (1_objem_3 : 5_objem_3)
```

4.3.5 Nejmenší a největší profit z transakcí

Dále z prvních pěti transakcí klienta vyčleníme tu, která přinesla největší a nejmenší profit společnosti.

Max_profit (největší profit)

```
MAX (1_profit_3 : 5_profit_3)
```

Min_profit (nejmenší profit)

```
MIN (1_profit_3 : 5_profit_3)
```

4.3.6 Průměrný profit a objem

Prumerny_profit (Průměrný profit z prvních pěti transakcí klienta)

```
1_5_profit_3 / Počet_tr_1_3
```

Prumerny_objem (Průměrný objem z prvních pěti transakcí klienta)

```
1_5_objem_3 / Počet_tr_1_3
```

4.3.7 Medián profitu a objemu

Median_profitu (Medián profitu z prvních pěti transakcí klienta)

```
MEDIAN(1_profit_3 : 5_profit_3)
```

Median_objemu (Medián objemu z prvních pěti transakcí klienta)

```
MEDIAN(1_objem_3 : 5_objem_3)
```

4.4 Aplikace BI nástrojů

V této části budou za pomoci softwaru od společnosti IBM - SPSS Modeler zkoumána data a souvislosti mezi nimi. První částí bude analýza klíčových proměnných, dále zkoumání právní formy klientů, jednorázových klientů a následné odhalení bonitních klientů na počátku smluvního vztahu včetně sestavení samotného modelu.

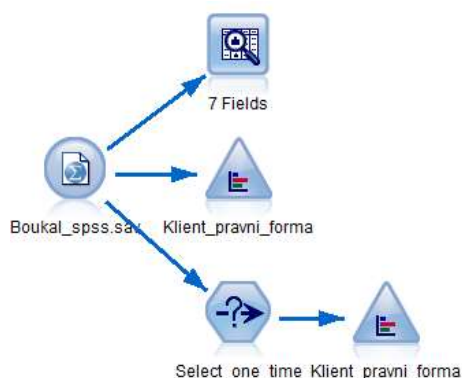
4.4.1 Analýza klíčových proměnných

Dalším postupem je zkoumání vybraných proměnných. Z velkého počtu jsou vybrány jenom ty, které nám následně budou sloužit k další analýze.

- *Klient_pravni_forma* - právní forma klienta
- *Pocet_tr_1_3* - počet transakcí uskutečněných za první 3 měsíce od podpisu smlouvy
- *Rozdeleni_klientu*- rozdělení klientů na základě jejich profitability
- *one_time* - klienti, kteří uskutečnili jednu nebo žádnou transakci
- *1_5_objem_3* - suma objemu pěti transakcí omezená na 3 měsíce od podpisu smlouvy
- *1_5_profit_3* - suma profitu pěti transakcí omezená na 3 měsíce od podpisu smlouvy
- *Prumer_transakci*- průměrná doba mezi prvními pěti transakcemi omezená na 3 měsíce od podpisu smlouvy

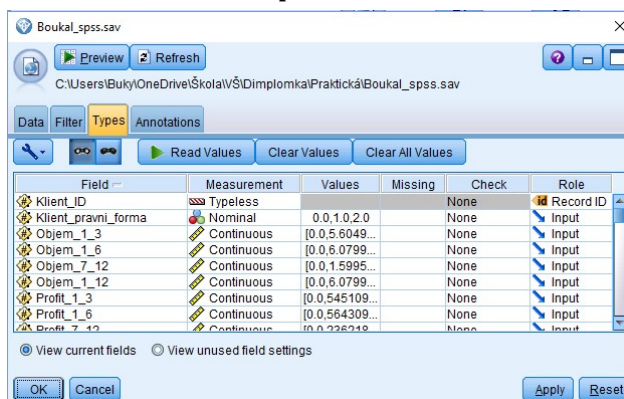
Výhodou SPSS Modeleru je snadná vizualizace postupu práce do tzv. streamu (proudu) dat, viz. obrázek č. 7. Prvním krokem je přiřazení datového souboru do streamu. V našem případě se jedná o soubor Boukal_spss.sav (jedná se o již převedený datový soubor z původního excelu s navíc vytvořenými proměnnými a rozdělením klientů podle jejich profitability z předešlých kroků 4.2 a 4.3).

Obrázek č. 7: Modeler Stream 1



Zdroj: Vlastní zpracování

Obrázek č. 8: Import datového souboru



Zdroj: Vlastní zpracování

Kromě logické adresy k souboru je zde možnost vyfiltrovat jen určité proměnné a v záložce „Types“ nastavit typ proměnných včetně jejich role (obrázek č. 8).

Prvně využijeme vestavěnou funkci **DATA AUDIT**, která umožňuje bližší pohled na jednotlivé proměnné. Na obrázku č. 9 je vidět výstup z této funkce. Vlevo vedle názvu proměnné je **histogram** vypovídající o rozložení dat. Dále můžeme vidět, o jaký typ dat se jedná (nominální/spojitá), minimální a maximální hodnotu, průměr, výběrovou odchylku, šikmost a medián.

U spojitých proměnných (**Measurement** = Continuous): počet transakcí, objem, profit a průměr transakcí. Z grafu je patrné **silné naklonění (šikmost) do leva**. Tento fakt potvrzuje i ukazatel „Skewness“, který při nulové hodnotě značí symetrické rozdělení. V našem případě se setkáváme i s hodnotou 20,9, to zna-

Obrázek č. 9: Data Audit

Field	Sample Graph	Measurement	Min	Max	Mean	Std. Dev	Skewness	Median	Valid
Klient_pravni_forma		Nominal	0.000	2.000	--	--	--	--	2589
Pocet_tr_1_3		Continuous	0.000	71.000	3.675	5.372	4.741	2.000	2589
Rozdeleni_klientu		Nominal	0.000	2.000	--	--	--	--	2589
one_time		Nominal	0.000	1.000	--	--	--	--	2589
@_1_5_objem_3		Continuous	0.000	109800000.000	1200183.891	3836910.897	15.155	374016.340	2589
@_1_5_profit_3		Continuous	0.000	225000.000	2108.648	6436.438	20.904	849.950	2589
@_Prumer_transakci		Continuous	0.000	91.000	17.771	16.451	1.747	14.000	2336

Zdroj: Vlastní zpracování

mená, že v datech se vyskytuje větší množství menších hodnot a menší množství velkých hodnot. V tomto konkrétním případě to poukazuje na fakt, že většina klientů uskutečňuje spíše méně drobných transakcí, které společnosti nepřinášejí významný zisk.

Nominální proměnné jsou v našem případě: *Klient_pravni_forma*, *Rozdeleni_klientu* a námi nově vytvořená proměnná *one_time*.

Hodnota **Min a Max** ukazuje minimální a maximální hodnotu vyskytující se v dané proměnné. U nominálních proměnných nám ovšem tento ukazatel moc neřekne. Právní forma obsahuje jen hodnoty 0;1 a 2, a tak minimální hodnotou je právě 0, která označuje klienty podnikající jako právnické osoby a maximální hodnotou je 2, která značí fyzické osoby ne podnikatelé.

U spojitých proměnných můžeme spolu s ukazatelem **Mean** (průměr), vyčíst, jak moc se liší maximální a průměrná hodnota. Například u počtu uskutečněných transakcí za tři měsíce je maximální počet transakcí 71, ale průměrný počet pouze 3,675. Největší objem transakce z prvních pěti transakcí omezených na tři měsíce od podpisu smlouvy je 109 800 000,- Kč, průměrný pak 1 200 183,- Kč. Největší profit z prvních pěti transakcí je pak 225 000,- Kč a průměrný 2 108,- Kč. U průměrné doby mezi pěti prvními uskutečněnými transakcemi je maximální hodnota 91 dní, průměr je 17,771 dní.

Ukazatel **Std. Dev** (Standard Deviaton) ukazuje **směrodatnou odchylku** dané proměnné. Směrodatná odchylka ukazuje, jak moc jsou si hodnoty dané proměnné podobné nebo rozdílné. Čím menší je hodnota tohoto ukazatele, tím více jsou si podobné.

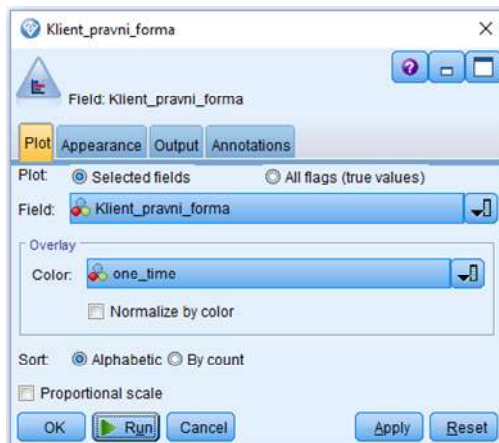
Medián je hodnota, jež dělí soubor hodnot na dvě stejně početné poloviny (50% hodnot je větších a 50% hodnot je menších než medián).

Ukazatel **Valid** značí počet použitelných hodnot v dané proměnné. Ve většině případů se jedná o hodnotu 2 589, tedy počet klientů v datovém souboru. Pouze u proměnné *Prumer_transakci* je tato hodnota menší (2 336) z důvodu chybějících hodnot.

4.4.2 Právní forma klientů

Blíže se podíváme na právní formu klientů. K přehledu hodnot znovu využijeme SPSS Modeler a konkrétně funkci **DISTRIBUTION**, která odhalí rozložení a počty klientů na základě jejich právní formy.

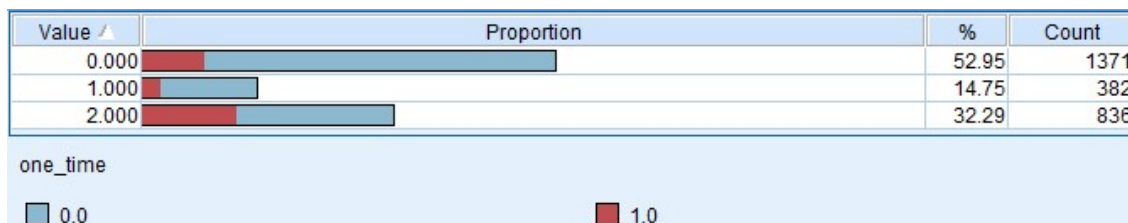
Obrázek č. 10: Nastavení funkce Distribution



Zdroj: Vlastní zpracování

Obrázek č. 10 ukazuje nastavení funkce distribution, zkoumáme proměnnou: *Klient_pravni_forma*. V nastavení lze ještě barevně oddělit jednorázové (one-time) klienty, aby byla možnost se podívat na zastoupení právě těchto klientů s ohledem na jejich právní formu.

Obrázek č. 11: Výstup funkce Distribution

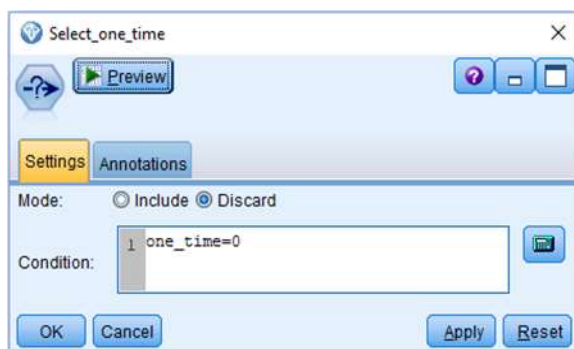


Zdroj: Vlastní zpracování

Na obrázku č. 11 je vidět výstup funkce distribution. Největší zastoupení klientů je 0 - právnické formy (52,95% - 1 371), následuje 2 - fyzická osoba ne podnikatel (32,29% - 836) a nejméně je fyzických osob podnikatelů (14,75% - 382).

Červená barva v grafu poté zobrazuje one-time klienty, tedy klienty kteří uskutečnili jednu nebo žádnou obchodní transakci. Jim se nyní budeme věnovat blíže.

Obrázek č. 12: Nastavení funkce Select



Zdroj: Vlastní zpracování

Pokud bychom si chtěli zobrazit pouze tyto klienty s ohledem na jejich právní formu, musíme využít funkce **SELECT**, která umožňuje nastavení specifických podmínek a zobrazení pouze data, která tuto podmínku splňují.

Ve funkci select nastavíme podmínku `one_time=0` se zaškrtnutým „discard“ (viz. obrázek č. 12). Tímto docílíme odfiltrování klientů, kteří mají v proměnné `one_time` hodnotu 0 (ti, kteří uskutečnili více než jednu obchodní transakci). Teď znovu použijeme funkci `distribution` k zobrazení hodnot.

Výstup je vidět na obrázku č. 13. Největší zastoupení `one_time` klientů je u právnické formy s označením 2 – fyzické osoby ne podnikatelé, jedná se o více jak polovinu (53,5% - 313 klientů), druhé nejvyšší zastoupení má 1 - právnické osoby (35,56% - 208 klientů) a nejméně je fyzických osob podnikatelů (10,94% - 64 klientů).

Obrázek č. 13: Výstup funkce Distribution

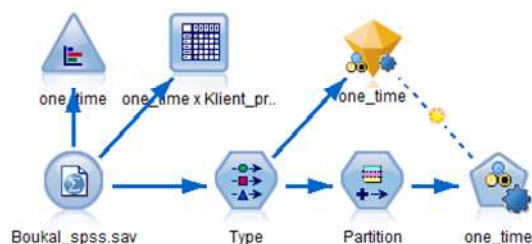
Value ▲	Proportion	%	Count
0.000		35.56	208
1.000		10.94	64
2.000		53.5	313

Zdroj: Vlastní zpracování

4.4.3 One-time klienti

Vzhledem k tomu, že klienti, kteří neuskuteční žádnou nebo pouze jednu obchodní transakci, jsou hrozbou pro společnost, podíváme se na ně blíže a zkusíme odhalit jejich společné znaky, aby je společnost mohla co nejdříve odhalit.

Obrázek č. 14: SPSS Modeler Stream 2



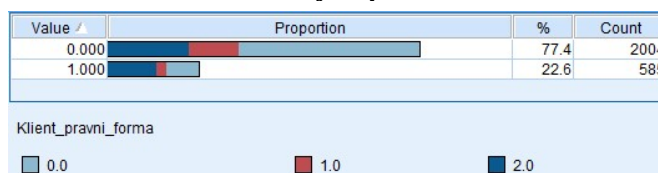
Zdroj: Vlastní zpracování

Na obrázku č. 14 je vidět postup práce (stream) z SPSS Modeleru.

Pro zkoumání se nejdříve podíváme na graf a tabulku rozdělení *one-time* zákazníků s ohledem na jejich právní formu. Využijeme znovu funkce **DISTRIBUTION**.

Graf navíc barevně odlišíme na základě právní formy klienta, abychom mohli opticky posoudit rozložení jednorázových klientů s ohledem na jejich právní formu.

Obrázek č. 15: Výstup funkce Distribution



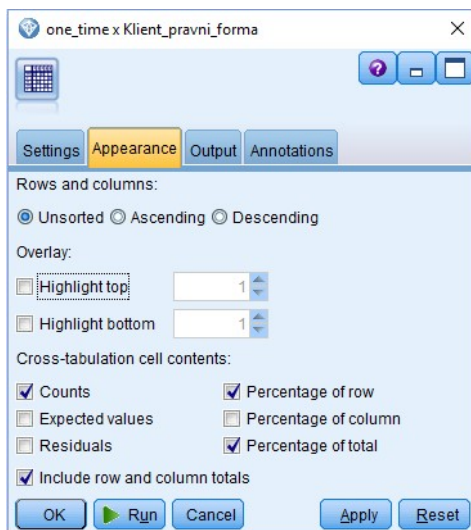
Zdroj: Vlastní zpracování

Výstup je poté vidět na obrázku č. 15. Klienti, kteří uskutečnili více jak jednu obchodní transakci, jsou označeni nulou a z celkového počtu 2 589 klientů se jedná o 77,4%, - 2 004 klientů.

Zbývají klienti 22,6% - 585, uskutečnili pouze jednu, nebo žádnou obchodní transakci.

K bližšímu pohledu využijeme funkce **MATRIX**, díky níž můžeme porovnat jednotlivé poměry a počty klientů s ohledem na jejich právní formu.

Obrázek č. 16: Nastavení funkce Matrix



Zdroj: Vlastní zpracování

Tato funkce nabízí různé možnosti zobrazení, viz. obrázek č. 16. Kromě klasického zobrazení dat do řádků a sloupců, je možnost zvolit v záložce *appearance* procentuální zastoupení v řádku, sloupci nebo celkové počty, očekávané hodnoty residuí, nebo zobrazení celkového počtu v řádcích i sloupcích.

Do řádku necháme zobrazit proměnnou *one-time* a do sloupců *Klient_pravni_forma*. V dalších možnostech zvolíme zobrazení procentuální zastoupení v řádku (*percentage of row*), celkové (*percentage of total*) a samozřejmě i jednotlivé počty (*counts*). Možnost celkového počtu v řádcích a sloupcích (*include row and column totals*) zvolíme také.

Obrázek č. 17: Výstup funkce Matrix

Klient_pravni_forma					
one_time		0.0	1.0	2.0	Total
0.0	Count	1163	318	523	2004
	Row %	58.034	15.868	26.098	100
	Total %	44.921	12.283	20.201	77.404
1.0	Count	208	64	313	585
	Row %	35.556	10.940	53.504	100
	Total %	8.034	2.472	12.090	22.596
Total	Count	1371	382	836	2589
	Row %	52.955	14.755	32.290	100
	Total %	52.955	14.755	32.290	100

Zdroj: Vlastní zpracování

Takto nastavená tabulka je vidět na obrázku č. 17. Ve skupině 0 (klienti s více jak dvěma transakcemi) je největší zastoupení skupiny 0 - právnických osob (58% - 1163), dále 2 - fyzických osob ne podnikatelů (26% - 523) a nejmenší zastoupení fyzických osob podnikatelů (15% - 318).

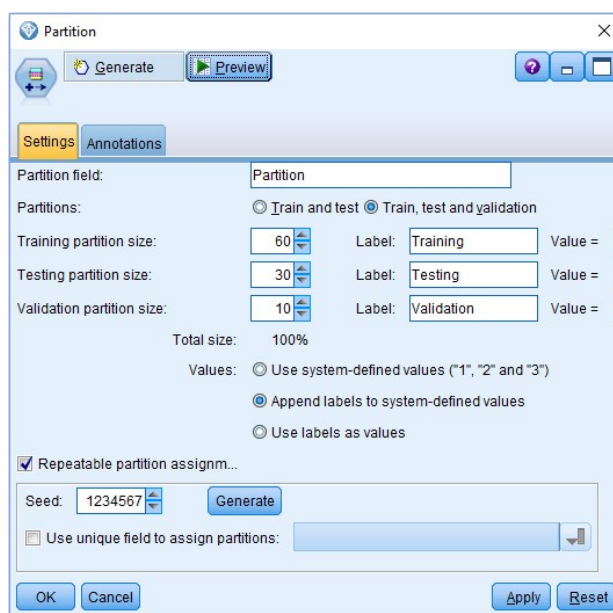
Klienti, kteří mají v proměnné *one-time* hodnotu 1 (jedna nebo žádná obchodní transakce), mají největší zastoupení fyzických osob ne podnikatelů (53% - 313), celkově se jedná o 12% ze všech klientů. Druhou nejpočetnější skupinou jsou právnické osoby (35% - 208) celkově se jedná o 8% všech klientů a nejméně je fyzických osob podnikatelů (11% - 64), celkově se jedná jen o 2,5% ze všech klientů.

Pro další zkoumání použijeme funkci **AUTO CLASSIFIER**. Tato funkce umožňuje porovnat více nástrojů klasifikace a na základě dat posoudit, která metoda se nejlépe hodí. Metoda klasifikace by měla být schopna odhalit společné znaky jednorázových klientů a pomoci společnosti tyto klienty dříve odhalit.

K přípravě dat pro tuto funkci použijeme znovu funkci **TYPE** a nově funkci **PARTITION**, která umožňuje rozdělení dat do více souborů (trénovací, testovací a validační). Toho lze následně využít v klasifikátoru pro zlepšení výsledků modelu, jak radí i E. Turban ve své knize *Business intelligence: a managerial approach*.

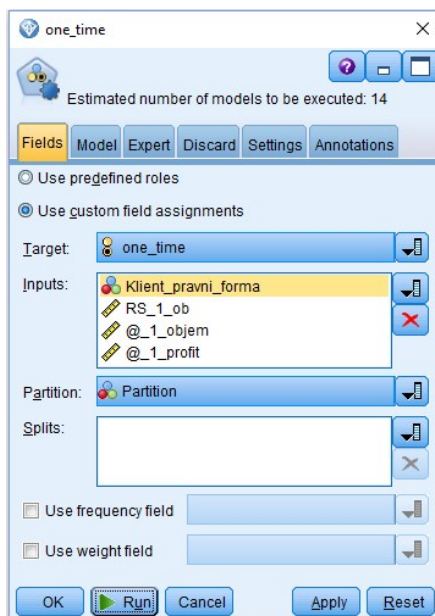
Obrázek č. 18 ukazuje nastavení funkce partition, rozdělení dat bude následující: 60% trénovacích, 30% testovacích a zbylých 10% validačních. Rozdělení dat je samozřejmě náhodné a tak by nemělo negativně ovlivnit výsledky klasifikace.

Obrázek č. 18: Nastavení funkce Partition



Zdroj: Vlastní zpracování

Obrázek č. 19: Nastavení funkce Auto Classifier



Zdroj: Vlastní zpracování

Na obrázku č. 19 je poté vidět **nastavení funkce auto classifier**. Jako cílovou (*target*) proměnnou zvolíme *one-time*. Jako vstupní (*inputs*) proměnné zvolíme: *Klient_pravni_forma*. Dále proměnné vypovídající o první transakci klienta, protože další informace o klientech, kteří uskutečnili maximálně jednu transakci, nemáme. Patří sem doba k uskutečnění první transakce od podpisu smlouvy-*RS_1_ob*, objem první transakce - *1_objem* a profit první transakce - *1_profit*. Do možnosti *partition* nastavíme stejně pojmenovanou proměnnou, která se vytvořila v předešlém kroku a rozděluje data do 3 souborů.

V záložce expert je možnost nastavení, které z klasifikačních modelů bude software porovnávat. Na výběr jsou metody klasifikačních stromů (*C5*, *Tree-AS*, *Chaid*, *C&R Tree*, *Random Trees*, *Quest*), neuronových sítí, logistické regrese, diskriminačních analýz a další.

Po spuštění auto classifieru nám SPSS zobrazí nejvhodnější modely pro daný typ dat a seřadí je podle přesnosti (*Overall Accuracy*) viz. obrázek č.19. Na výběr software vybral tři modely s přesností kolem 78%, což lze považovat za uspokojivý výsledek.

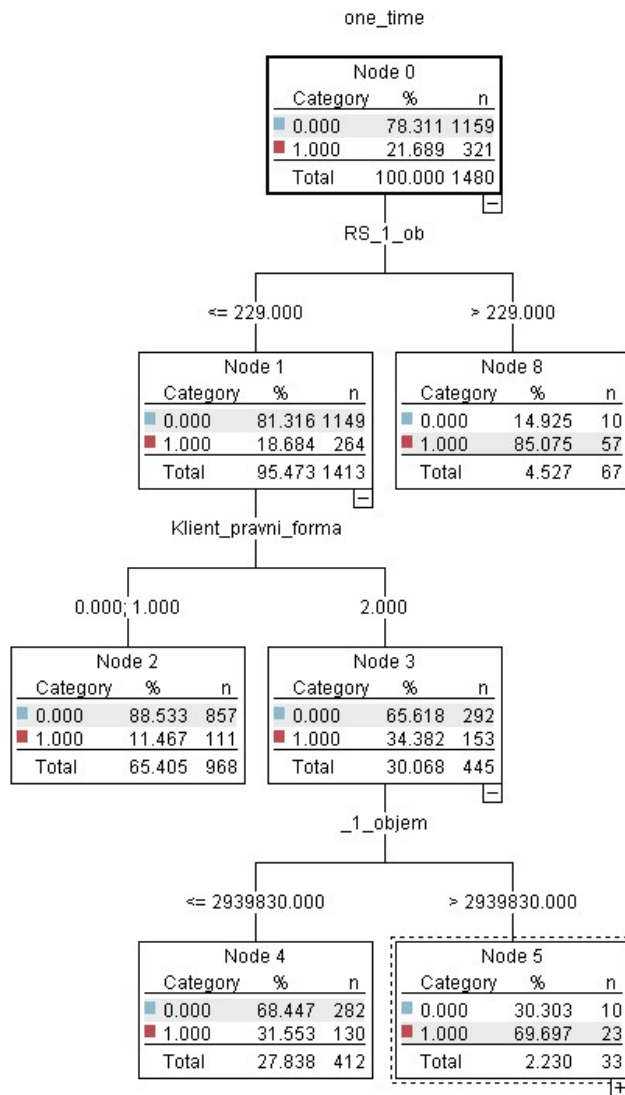
Jako nejpresnější vybral metodu rozhodovacích stromů *C5 1*, jejíž výstup je zobrazen na obrázku č. 20.

Obrázek č. 20: Výstup Auto Classifieru

Use?	Graph	Model	Build Time (mins)	Max Profit	Max Profit Occurs in (%)	Lift (Top 30...	Overall Accuracy (%)	No. Fields Used	Area Under Curve
<input checked="" type="checkbox"/>		C5 1	< 1	111,86	5	1,702	78,521	3	0,68
<input checked="" type="checkbox"/>		C&R Tree 1	< 1	122,222	7	1,798	78,169	4	0,697
<input checked="" type="checkbox"/>		SVM 1	< 1	110,0	4	1,515	78,052	4	0,661

Zdroj: Vlastní zpracování

Obrázek č. 21: Rozhodovací strom C5



Zdroj: Vlastní zpracování

Na obrázku č. 21 je vidět výstup metody C5 1, jedná se o statistickou metodu rozhodovacího stromu, která na každé větvi reprezentuje jedno dílčí rozdělení.

NODE 0 označuje **počáteční stav** proměnné – 1 480 klientů, z toho 1 159 s hodnotou proměnné one-time 0 a 321 s hodnotou 1 (klienti s jednou nebo žádnou transakcí).

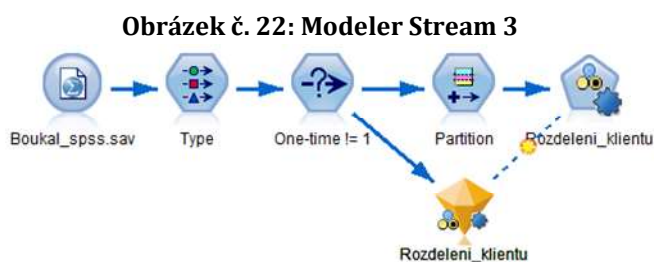
První větev rozděluje klienty podle **doby uskutečnění první transakce** od podpisu smlouvy (*RS_1_ob*). **Pokud je tato hodnota vyšší než 229 dní, jedná se s jistotou 85,1% o one-time klienta.** Pokud klient uskuteční první obchod dříve než za 229 dní, je stále 18,7% možnost, že další transakci neuskuteční.

Další větev se týká pouze většího souboru z předešlého rozvětvení, v našem případě NODE1. Klienty, kteří do tohoto souboru spadají (1 413), model dále rozděluje podle právní formy (*Klient_pravni_forma*). **Větší pravděpodobnost k tomu být one-time klientem mají klienti s právní formou 2 (fyzická osoba ne podnikatel),** zde se jedná o pravděpodobnost 34,4%, naopak u právnických osob a fyzických osob podnikatelů se jedná pouze 11,5%.

Poslední větev se týká NODE 3 (445 klientů), která je rozdělena podle objemu první uskutečněné transakce (*1_objem*). **Pokud je první uskutečněná transakce větší než 2 939 830 Kč, je také větší pravděpodobnost že klient další transakci neuskuteční a bude se tak jednat o one-time klienta (69,7%),** v opačném případě je tato pravděpodobnost méně než poloviční - 31,6%.

4.5 Odhalení bonitního klienta na začátku smluvního vztahu

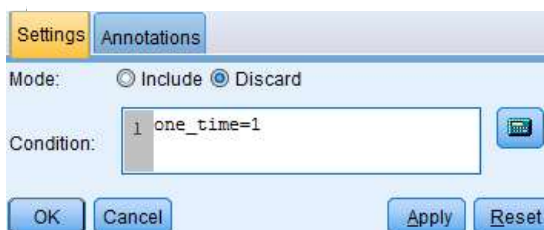
Na začátku jsme si rozdělili klienty podle jejich ziskovosti na 0 - *ne profitabilní*, 1 - *krátkodobě profitabilní* a 2 - *dlouhodobě profitabilní*. Dalším krokem je aplikace data mining a business intelligence nástrojů k odhalení možných spojitostí za účelem odhadnutí tohoto rozdělení na začátku smluvního vztahu. Požadavkem bylo odhalit klienta ve zkušební době, která je omezena na 3 měsíce od podpisu smlouvy a na prvních 5 uskutečněných transakcích. Tomu nám odpovídají i data, která jsme obdrželi v 3. datovém souboru.



Zdroj: Vlastní zpracování

Na obrázku č. 22 je vidět postup práce v SPSS Modeleru. Nejdříve využijeme funkce **TYPE** pro načtení proměnných a zvolení jejich rolí. Jak jsme již dříve zjistili, v datech se vyskytuje velké množství klientů, kteří uskutečnili jednu nebo žádnou obchodní transakci a mohli nepříznivě ovlivnit výsledek modelu. Pomocí funkce **SELECT** tyto klienty vyřadíme, aby nám zůstali pouze klienti, s více jak jednou obchodní transakcí (obrázek č. 23).

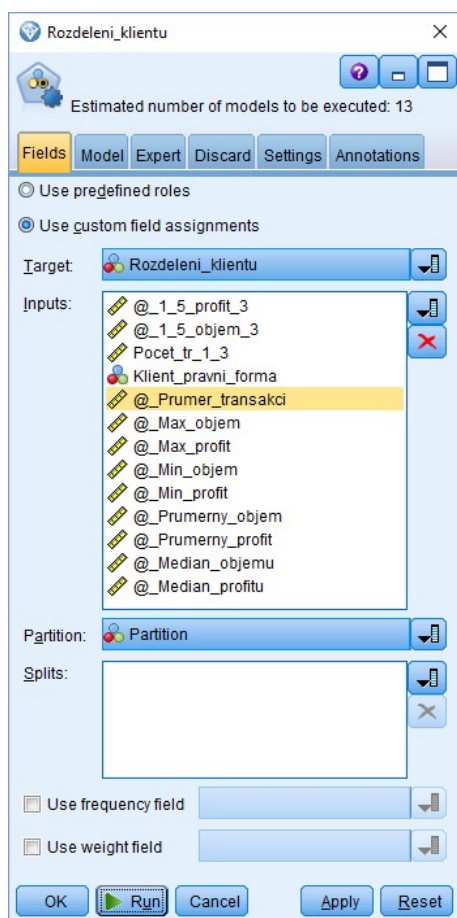
Obrázek č. 23: Funkce Select



Zdroj: Vlastní zpracování

Pro lepší věrohodnost modelu znovu použijeme funkci **PARTITION** a rozdělíme data do 3 souborů (trénovací, testovací a validační).

Obrázek č. 24: Nastavení Auto classifieru 2



Zdroj: Vlastní zpracování

Obrázek č. 25: Výstup z Auto classifieru 2

Use?	Graph	Model	Build Time (mins)	Overall Accuracy (%)	No. Fields Used
<input checked="" type="checkbox"/>		CHAID 1	< 1	66,195	6
<input checked="" type="checkbox"/>		C&R Tree 1	< 1	65,252	12
<input checked="" type="checkbox"/>		Quest 1	< 1	64,465	6

Zdroj: Vlastní zpracování

Dále použijeme funkci **AUTO CLASSIFIER**, zde je pro nás zkoumanou (cílovou) proměnnou *Rozdeleni_klientu* a jako vstup všechny proměnné, které by nám mohly pomoci odhalit spojitosti na začátku smluvního vztahu:

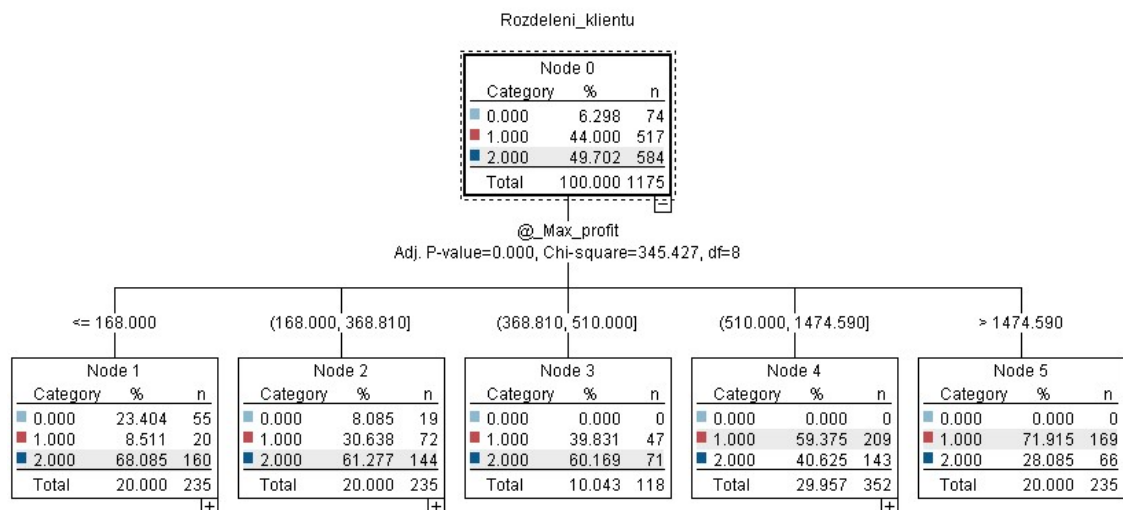
1_5_profit_3, *1_5_objem_3*, *Pocet_tr_1_3*, *Klient_pravni_forma* a *Prumer_transakci*.

SPSS jako nejvhodnější modely vybral *CHAID*, *C&R Tree* a *Quest*. Nejvyšší přesnosti dosahuje *CHAID* s 66,195% (obrázek č. 24).

Výstupem modelu *CHAID* je rozhodovací strom. Model je sestaven až do hloubky 4, kde nejmenší uzly čítají méně než 20 klientů. Vzhledem k obsáhlosti celého modelu i samotného stromu bude zobrazena jenom část.

Model začíná v **NODE 0** s 1 175 klienty. První rozvětvení je na základě proměnné **Max_profit** (maximální profit z prvních pěti transakcí za 3 měsíce

Obrázek č. 26: Výstup modelu Chaid



Zdroj: Vlastní zpracování

od podpisu smlouvy) a dělí se mezi 5 dalších uzlů (viz. obrázek č. 26).

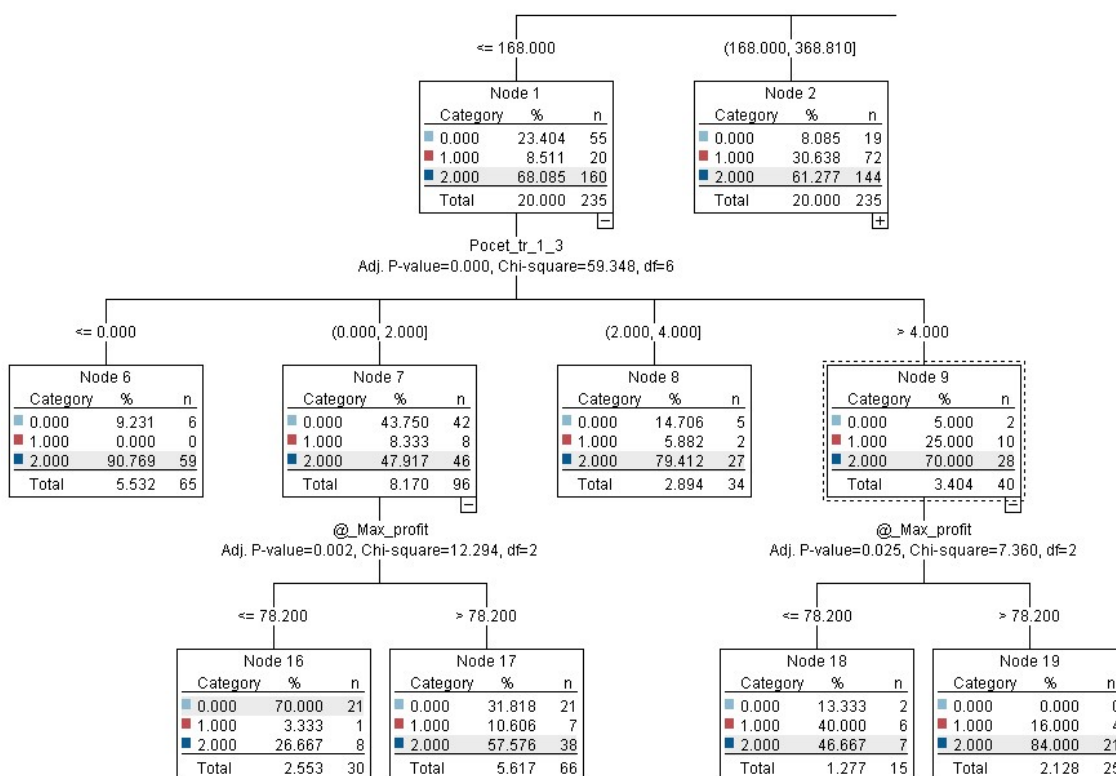
Z prvního rozvětvení můžeme vyvodit následující závěry:

NODE 1 a *NODE 2* jsou jediné uzly, kde se vyskytují *ne profitabilní* klienti (*Rozdeleni_klienta* = 0). Pokud je tedy **Max_profit** větší než **368,81 Kč**, nejedná se o **ne profitabilního** klienta.

Procentuální pravděpodobnost výskytu *krátkodobě profitabilního* klienta je nejnižší v *NODE 1* (8,511%), naopak u *dlouhodobě profitabilního* klienta se jedná o pravděpodobnost 68,085%. Se vzrůstající hodnotou proměnné *Max_profit* tato pravděpodobnost vzrůstá na úkor *dlouhodobě profitabilních* klientů. V posledním uzlu na této úrovni rozvětvení (*NODE 5*) je pravděpodobnost výskytu *krátkodobě profitabilního* klienta již 71,915%, ale u *dlouhodobě profitabilního* klienta jen 28,085%.

Čím vyšší je Max_profit, tím je větší pravděpodobnost výskytu krátkodobě profitabilního klienta (*Rozdeleni_klienta* = 1).

Obrázek č. 27: Výstup Chaid 2



Zdroj: Vlastní zpracování

Obrázek č. 27 zobrazuje rozvětvení **NODE 1** do dalších uzlů. První dělení nastává na základě proměnné **Pocet_tr_1_3**. Pokud je počet transakcí za první tři měsíce od podpisu smlouvy roven 0 (**NODE 6**), je 90,769% pravděpodobnost, že se bude jednat o *dlouhodobě profitabilního* klienta, zbylých 9,231% vyjadřuje pravděpodobnost na *ne profitabilního* klienta.

Pokud je **Pocet_tr_1_3** větší než 4 a **Max_profit** větší než 78,2 Kč a zároveň menší než 168 Kč, nejedná se o *ne profitabilního* klienta, ale s pravděpodobností 84% o klienta *dlouhodobě profitabilního* (**NODE 19**).

5 Shrnutí výsledků

V praktické části diplomové práce byla zkoumána data z reálné finanční instituce **AKCENTA CZ, a.s.**. K dispozici byly **informace o 2 594 klientech** a jejich uskutečněných transakcích. Data byla rozdělena do tří datových sad. První sada sloužila k identifikaci klienta, druhá sada poskytovala informace o prvních pěti uskutečněných transakcích a třetí sada byla od druhé omezena na zkušební dobu tří měsíců od podpisu smlouvy.

Klienty bylo nejprve nutné rozdělit podle jejich profitability na *ne profitabilní, krátkodobě a dlouhodobě profitabilní*. Při prohlížení dat bylo odhaleno velké množství klientů, kteří uskutečnili jednu nebo dokonce žádnou obchodní transakci v horizontu jednoho roku od podpisu smlouvy. Tito klienti byli následně vyčleněni a v další části blíže zkoumáni. Celkem se jednalo téměř o čtvrtinu všech klientů, 22,6% - 585 klientů.

Největší riziko žádné nebo jedné uskutečněné transakce mají z hlediska právní formy fyzické osoby ne podnikatelé (53%), dále právní osoby (35%) a nejmenší riziko představují fyzické osoby podnikatelé (11%). Při odhalování takto neaktivního klienta je důležitým prvkem **doba, za kterou uskutečnil první transakci. Pokud tak neučiní do 229 dnů od podpisu smlouvy, je 85,1% pravděpodobnost, že v prvním roce neuskuteční více jak jednu obchodní transakci.** Zajímavé je zjištění, že **pokud bude první uskutečněná transakce vyšší než 2 936 830 Kč, lze na 69,7% očekávat, že tato klientova transakce bude i jeho poslední.**

Blíže byli zkoumáni i klienti s ohledem na jejich právní formu. **Největší zastoupení je právnických osob (52,95% - 1 371), dále fyzických osob ne podnikatelů (32,29% - 836) a nejméně je fyzických osob podnikatelů (14,75% - 382).**

Hlavním cílem diplomové práce bylo odhalit bonitního klienta na počátku smluvního vztahu. Požadavkem bylo omezení dat na zkušební dobu, která spočívá ve třech měsících od podpisu smlouvy a v prvních pěti uskutečněných transakcích. Vzhledem k tomu, že ne všichni klienti uskutečnili pět transakcí (někteří neuskutečnili ve zkušební době žádnou), bylo nutné některá **data sumarizo-**

vat a zprůměrovat, aby mohla lépe charakterizovat klientovo chování. Celkem bylo vytvořeno 12 nových proměnných, které byly následně využity k sestavení modelu.

K tomu byl využit **statistický software SPSS Modeler**. Výstupem je rozhodovací strom, který na základě dat ze zkušební doby je schopen předpovídat do jaké ze skupin klient bude klient patřit (ne profitabilní, krátkodobě profitabilní, dlouhodobě profitabilní).

Model CHAID, který dosáhl přesnosti 66,2%, vybral jako proměnnou, která může nejlépe rozlišit budoucí profitabilitu klienta, maximální profit z prvních pěti transakcí klienta. **Pokud bude maximální profit z transakce větší než 368,81 Kč, nebude se jednat o ne profitabilního klienta.** Naopak čím vyšší je tato hodnota, tím je větší pravděpodobnost, že klient bude pouze krátkodobě profitabilní na úkor výskytu dlouhodobě profitabilního klienta.

Klíčem k odhalení ne profitabilního klienta je maximální profit z transakce. Pokud je menší než 78,2 Kč a počet transakcí uskutečněných za první tři měsíce od podpisu smlouvy je roven 1 nebo 2, je zde 70% pravděpodobnost, že klient bude ne profitabilní.

Problémem v budoucnu je však relativně vysoký výskyt profitabilních klientů v situacích, které by spíše nasvědčovaly o opaku. **S pravděpodobností 90,1% bude klient, který neuskuteční žádnou obchodní transakci v horizontu tří měsíců od podpisu smlouvy dlouhodobě profitabilní.**

Z tohoto důvodu je rozřazení klientů ve zkušební době, která je omezená na tři měsíce od podpisu smlouvy a prvních pěti uskutečněných transakcí velice riziková a může dojít k nesprávnému zařazení dlouhodobě profitabilního klienta mezi ne profitabilní.

K lepšímu a spolehlivějšímu přiřazení klientů ve zkušební době by bylo zapotřebí více informací a dat, které by například vypovídaly o prosperitě klienta a mezinárodním působení (obor podnikání, obrat, velikost, zisk, export, import). Tyto a další informace je ovšem složité získat ke každému klientovi a především nesplňují podmínku anonymity, která byla požadavkem společnosti pro vypracování praktické části diplomové práce.

6 Závěr

Hlavním cílem diplomové práce bylo odhalení bonitního klienta na počátku smluvního vztahu. Tento cíl byl blíže specifikován na zkušební dobu třech měsíců a požadavek rozdělení klientů na ne profitabilní, krátkodobě a dlouhodobě profitabilní. Cíle bylo dosaženo díky sestavení klasifikačního modelu softwarem SPSS Modeler, jehož výstupem je rozhodovací strom. Ten umožňuje odhalení společných znaků jednotlivých skupin a predikovat tak ve zkušební době budoucí profitabilitu či ne profitabilitu klienta. Model dosáhl přesnosti 66,2%, to lze, na ekonomický typ modelu, považovat za přijatelné. Část dat byla navíc vyčleněna na testování a validaci modelu.

Díličními cíli bylo odhalení podezřelých nebo rizikových klientů pro společnost. Bylo objeveno relativně velké množství klientů, kteří uskutečnili jednu nebo dokonce žádnou obchodní transakci. Celkem se jednalo o 22,6% - 585 klientů. Tito klienti byli označeni a blíže zkoumáni, znovu byl sestaven klasifikační model v SPSS Modeleru.

Společnost na základě sestavených modelů je schopna alespoň většinu klientů roztrždit ve zkušební době a přizpůsobit jim svoji obchodní strategii. Rizikem jejich brzkého rozřazení je zařazení v budoucnu profitabilních klientů mezi ne profitabilní a jejich následné odrazení od další spolupráce. Z tohoto důvodu je vhodné posouzení i dalších faktorů, které vzhledem k anonymitě dat nemohly být použity (velikost klientova podnikání, mezinárodní zapojení do obchodu, osobní zkušenosti apod.).

Modely by bylo vhodné po čase znovu přepočítat a použít aktuální data klientů, dlouhodobé použití stejného modelu není vhodné vzhledem k měnícím se podmínkám, které mohou ovlivnit chování klientů (růst/recese ekonomiky, výkyvy kurzů apod.).

7 Seznam použité literatury

[1] BIERE, Mike. *Business intelligence for the enterprise*. 2nd print. Upper Saddle River, N.J.: Prentice Hall PTR, 2003. ISBN 978-013-1413-030.

[2] *Česká národní banka* [online]. 2016 [cit. 2016-01-19]. Dostupné z: <https://www.cnb.cz>

[3] GÁLA, Libor, Jan POUR a Zuzana ŠEDIVÁ. *Podniková informatika. 2.*, přeprac. a aktualiz. vyd. Praha: Grada, 2009. Expert (Grada). ISBN 978-80-247-2615-1.

[4] JÍLEK, Josef. *Finanční trhy*. Vyd. 1. Praha: Grada, 1997, 527 s. ISBN 80-716-9453-3.

[5] JÍLEK, Josef. *Termínové a opční obchody*. Vyd. 1. Praha: Grada, 1995, 286 s. ISBN 80-716-9183-6.

[6] MUSÍLEK, Petr. *Trhy cenných papírů*. Vyd. 1. Praha: Ekopress, 2002, 459 s. ISBN 80-861-1955-6.

[7] PIVIDORI, Milton, Georgina STEGMAYER a Diego H. MILONE. Diversity kontrol for improving the analysis of consensus clustering. *Information Sciences*. 2016, (361-362), 120-134.

[8] POLOUČEK, Stanislav. *Peníze, banky, finanční trhy*. Vyd. 1. V Praze: C.H. Beck, 2009, xvii, 415 s. Beckovy ekonomické učebnice. ISBN 978-80-7400-152-9.

[9] Registr ekonomických subjektů. *Český statistický úřad*. [online]. 12.8.2016 [cit. 2016-08-12]. Dostupné z: http://apl.czso.cz/irsw/detail.jsp?prajed_id=723644

[10] REVENDA, Zbyněk. *Peněžní ekonomie a bankovníctví*. 5., aktualiz. vyd. Praha: Management Press, 2012, 423 s. ISBN 978-80-7261-240-6.

[11] SKALSKÁ, Hana. *Data mining a klasifikační modely*. Vyd. 1. Hradec Králové: Gaudeamus, 2010. ISBN 978-80-7435-088-7.

[12] The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things. EMC Corporation.[online]. Duben 2014 [cit. 2016-03-25]. Dostupné z: <http://www.emc.com/leadership/digital-universe/2014iview/index.htm>

[13] TUFFERY, Stéphane. *Data mining and statistics for decision making*. Hoboken, NJ.: Wiley, 2011. Wiley series in computational statistics. ISBN 978-0-470-68829-8

[14] TURBAN Efraim. [ET AL.]. *Business intelligence: A managerial approach*. 2nd ed., International ed. Boston: Prentice Hall, 2011. ISBN 01-324-7882-X.

[15] TURNER, Dawn M. *What is Venture Management?* [online]. In: . 2016 [cit. 2016-07-19]. Dostupné z: <http://www.ventureskies.com/blog/what-is-venture-management>

[16] Veřejný rejstřík a Sběrka listin. *Ministerstvo spravedlnosti České republiky*. [online]. 13.8.2016 [cit. 2016-08-13]. Dostupné z: <https://or.justice.cz/ias/ui/rejstřík-firma.vysledky?subjektId=286916&typ=PLATNY>

[17] Zahraniční měny a platby – AKCENTA CZ. *Akcenta*. [online]. 13.8.2016 [cit. 2016-08-13]. Dostupné z: <http://www.akcentacz.cz>

[18] Zákon č. 21/1992 Sb. o bankách. Dostupný z: <https://portal.gov.cz/app/zakony/zakonPar.jsp?idBiblio=39677&nr=21~2F1992&rp=15#local-content>

8 Seznam obrázků, grafů a tabulek

Obrázek č. 1: Množství dat vhodných k dalšímu zpracování	3
Obrázek č. 2: Rozhodovací strom	7
Obrázek č. 3: Shlukové metody	8
Obrázek č. 4: CRISP-DM proces	9
Obrázek č. 5: Kroky přípravy dat pro DM.....	10
Obrázek č. 6: SEMMA Data Mining proces	12
Obrázek č. 7: Modeler Stream 1	33
Obrázek č. 8: Import datového souboru	34
Obrázek č. 9: Data Audit	34
Obrázek č. 10: Nastavení funkce Distribution	36
Obrázek č. 11: Výstup funkce Distribution	36
Obrázek č. 12: Nastavení funkce Select.....	37
Obrázek č. 13: Výstup funkce Distribution	37
Obrázek č. 14: SPSS Modeler Stream 2.....	38
Obrázek č. 15: Výstup funkce Distribution	38
Obrázek č. 16: Nastavení funkce Matrix.....	39
Obrázek č. 17: Výstup funkce Matrix.....	39
Obrázek č. 18: Nastavení funkce Partition.....	40
Obrázek č. 19: Nastavení funkce Auto Classifier	41
Obrázek č. 20: Výstup Auto Classifieru.....	42
Obrázek č. 21: Rozhodovací strom C5	42
Obrázek č. 22: Modeler Stream 3.....	44
Obrázek č. 23: Funkce Select	44
Obrázek č. 24: Nastavení Auto classifieru 2	45
Obrázek č. 25: Výstup z Auto classifieru 2	45
Obrázek č. 26: Výstup modelu Chaid.....	46
Obrázek č. 27: Výstup Chaid 2	47
Graf č. 1: Počty klientů při přerozdělování do skupin	29
Graf č. 2: Počet jednorázových klientů	30

Tabulka č. 1: Definiční charakteristiky investičních zprostředkovatelů a komerčních bank.....	20
Tabulka č. 2: První rozdělení (počty klientů)	27
Tabulka č. 3: Průměr a střední hodnota počtu transakcí	27
Tabulka č. 4: Druhé rozdělení (počty klientů).....	28
Tabulka č. 5: Konečné rozdělení (počty klientů).....	29