

Vysoká škola ekonomická v Praze

Fakulta informatiky a statistiky

Studijní program: Kvantitativní metody v ekonomice

Studijní obor: Statistické metody v ekonomii



**APLIKACE STATISTICKÝCH METOD PŘI VÝPOČTU CELOŽIVOTNÍ
HODNOTY ZÁKAZNÍKA**

Bakalářská práce

Autorka bakalářské práce: Katsiaryna Samuseva

Vedoucí bakalářské práce: Ing. Lenka Vraná

Akademický rok 2016/2017

Prohlášení

Prohlašuji, že jsem bakalářskou práci zpracovala samostatně a že jsem uvedla všechny použité prameny a literaturu, ze kterých jsem čerpala.

V Praze dne 05. 01. 2017

.....
podpis

Poděkování

Mé poděkování patří Ing. Lence Vrané za odborné vedení, trpělivost a ochotu, kterou mi v průběhu zpracování bakalářské práce věnovala.

Abstrakt

Bakalářská práce se věnuje statistickým metodám pro řízení vztahů se zákazníky a různým přístupům k odhadu celoživotní hodnoty zákazníka pro společnost v dlouhodobém horizontu. Jde o empirický typ práce, kde teoretické předpoklady otestujeme v praxi na reálných datech. V průběhu práce zkoumáme a měříme klientské chování s cílem získat kvantitativní výstupy pro přizpůsobení marketingové strategie, efektivní alokaci zdrojů a výběr cílových zákazníků.

Na základě chování v minulosti a současnosti budeme modelovat očekávaný budoucí stav nákupního chování a s tím spojené vymezení nejziskovější cílové skupiny pro kvalitnější aplikaci marketingových nástrojů. Mezi hlavní cíle práce patří popis různých přístupů k modelování CLV, aplikace modelu Pareto/NBD na reálná data společnosti „Práce pro Vás“ a zhodnocení využitelnosti výsledků predikování CLV v praxi. V průběhu práce budou řešeny i následující úkoly: uvést základní výhody a nevýhody jednotlivých způsobů modelování CLV, vyhodnotit kvalitu modelu Pareto/NBD a stanovit další náměty při zkoumání této problematiky. Existují různé přístupy k odhadu CLV, které jsou popsány v teoretické části. Praktická část práce je zaměřena především na aplikaci pravděpodobnostního modelu Pareto/NBD. Pomocí modelu Pareto/NBD manažeři získávají informace o budoucí aktivitě zákazníka, tj. očekávaný počet transakcí na základě jeho skutečného chování a pravděpodobnost, že zákazník bude stále aktivní ve sledovaném horizontu. V současné realitě „přetíženého trhu“ téměř ve všech odvětvích podnikání musí společnosti spoléhat na kvantitativní podstatu marketingu a vybírat si cílové zákazníky, na které zaměří své marketingové úsilí. Přesně této potřebě odpovídá model Pareto/NBD, pomocí kterého odhadneme celoživotní hodnotu zákazníka pro strategické rozhodování a řízení vztahů se zákazníky. Závěrem se nám podaří vyřešit dané problémy a splníme tím všechny stanovené cíle.

Klíčová slova: celoživotní hodnota zákazníka, CLV, Pareto/NBD, CRM, pravděpodobnostní modely

Abstract

This bachelor thesis pay attention on statistical methods managing relationship with customers and different approaches of estimation customer lifetime value for the company in long time period. It's an empirical type of bachelor thesis, where we will test the theoretical assumptions on the real data. During this paper we research and measure customer's behavior in order to obtain quantitative outputs to adaptation marketing strategy, efficient allocation of resources and choosing key clients. On the basis of past and present behavior we simulate expected in future level of purchasing behavior and thus try to determine the most profitable target groups for high-quality application of marketing tools. Main purposes of thesis are describing various attitudes to modeling CLV, applying the model Pareto/NBD on real date of the company "Práce pro Vás ", to evaluate forecast applicability in practice. Following tasks will be resolved in this paper: to introduce advantages and disadvantages of each modeling CLV approach, to evaluate quality of Pareto/NBD model and to determine the advanced extensions of the research in this issue. There are various attitudes to estimation of CLV, which are described in theoretical chapters, but practical part of this paper focuses primarily on application of probability model Pareto/NBD. With the help of Pareto/NBD managers can get the information about future customer's activity, i.e. expected number of transactions from real behavior a probability that customer will be active in observation period. In nowadays reality of „overloaded market “, almost in every business sector, every company should rely on quantitative base of marketing and choose target clients, on which to focus their marketing efforts. Exactly this need matches with described and applied in this bachelor thesis model Pareto/NBD, with it help we will estimate customer lifetime value for strategic decisions and managing relationship with customers. Finally, we will resolve continuous tasks and will achieve all the determined goals.

Key words: Customer Lifetime Value, CLV, Pareto/NBD, CRM, Probability Models

Obsah

1. Úvod.....	1
1.1 Cíle práce.....	1
1.2 Struktura práce.....	2
2. Celoživotní hodnota zákazníka CLV.....	4
2.1 Popis konceptu CLV a příklad jeho výpočtu.....	4
2.2 Způsoby modelování CLV.....	7
2.3 Koncept řízení vztahů se zákazníky CRM.....	17
2.4 Popis modelu Pareto/NBD.....	19
3. Aplikace Pareto/NBD modelu na reálná data společnosti.....	27
3.1 Krátký popis analyzované společnosti.....	27
3.2 Popis řešeného problému.....	28
3.3 Popis a příprava reálných dat.....	29
3.4 Aplikace modelu Pareto/NBD na reálná data pomocí R.....	32
3.5 Zhodnocení modelu Pareto/NBD.....	43
3.6 Finanční analýza efektivity modelu Pareto/NBD.....	45
Závěr.....	50
Seznam literatury.....	53
Seznam obrázků a tabulek.....	57

1. Úvod

Bakalářská práce je zaměřena na aplikace statistických metod pro výpočet celoživotní hodnoty klientů a řízení vztahů se zákazníky pomocí predikce nákupního chování. Z důvodu globalizace a neustálé rostoucí konkurence je toto téma stále aktuální a představuje velký význam pro strategické rozhodování, efektivní alokaci zdrojů a výběr cílových zákazníků. Důvod a motivace pro zkoumání této problematiky spočívá v potřebě kvantifikovat výstupy marketingových aktivit firmy a prokázat, že celoživotní hodnota klienta představuje velký zájem pro každou společnost a je nedílnou součástí strategického rozhodování. Pomocí odhadu celoživotní hodnoty manažeři společností získávají úplný přehled očekávaného chování zákazníka, pomocí kterého mají i informaci o budoucích příjmech společnosti.

1.1 Cíle práce

Mezi hlavní cíle bakalářské práce patří:

- popsat různé přístupy k modelování CLV;
- aplikovat model Pareto/NBD na reálná data společnosti „Práce pro Vás“;
- zhodnotit využitelnost výsledků predikování CLV v praxi.

Průběžné úkoly práce:

- uvést základní výhody a nevýhody jednotlivých způsobů modelování CLV;
- vyhodnotit kvalitu modelu Pareto/NBD;
- stanovit další náměty při zkoumání této problematiky.

Vzhledem k definovaným cílům a záměrům můžeme zařadit bakalářskou práci k empirickému typu šetření. Hlavní funkce empirického výzkumu je ověření pravdivosti domněnek a hypotéz, které jsou výsledkem teoretického myšlení. Práce obsahuje teoretické předpoklady, které pak otestujeme v praxi na reálných datech. Podle základů empirického výzkumu pracujeme vždy s konkrétními daty a pomocí exaktních statistických nebo matematických metod dospíváme ke konkrétním výsledkům, které pak srovnáváme s teoretickými předpoklady. Základní metody empirického šetření jsou pozorování, měření a experiment. Jak uvidíme v subkapitole 1.2, která se věnuje struktuře bakalářské práce, náš případ odpovídá stanovenému typu.

Z našich cílů jasně vyplývají způsoby jejich dosažení: budeme zkoumat zákaznické chování na základě statistických metod a pravděpodobnostního modelu. Poté využijeme výstupy

modelu Pareto/NBD a odhadneme celoživotní hodnotu zákazníka na příkladu dat společnosti „Práce pro Vás“ jak na individuální úrovni, tak i pro celý datový soubor.

1.2 Struktura práce

Daná bakalářská práce se skládá z pěti kapitol. První kapitola slouží jako úvod, ve které jsme vymezili aktuální téma a vyjmenovali důvody pro zkoumání této problematiky. Na základě definování cílů práce a způsobu jejich dosažení jsme zařadili práci mezi empirické typy prací.

Druhá kapitola je určena zejména pro vymezení pojmu a podrobný popis podstaty CLV. Zároveň vyjmenujeme různé složky marketingových aktivit a ukážeme na příkladech, jak jednotlivé složky ovlivňují celoživotní hodnotu zákazníka. Dále charakterizujeme šest různých modelů a přístupů, které se používají k odhadu CLV, zmíníme výhody a nevýhody jednotlivých způsobů modelování a rozdíly mezi nimi. Poté vybereme konkrétní model pro aplikaci na reálných datech společnosti. Ve druhé kapitole také uvedeme současný stav poznání a provedeme rešerši zdrojů, která nám pomůže získat informace o literatuře, kde autoři zkoumali zákaznické chování, aplikovali modely a snažili se predikovat CLV pomocí různých statistických metod. Pokračujeme popisem nedílné součásti řízení vztahů se zákazníky – CRM. Tento koncept je aktuálním a zásadním nástrojem pro naše účely. Komplexní přístup umožňuje firmě transformovat CRM pomocí CLV na kvalitní a výkonný proces řízení vztahů se zákazníky a získat nepopíratelnou konkurenční výhodu. Pak následuje model Pareto/NBD, který jsme si vybrali pro predikci CLV a také zmíníme i několik variant rozšíření tohoto modelu. Pareto/NBD řeší následující úkoly: zjišťuje pravděpodobnost aktivity zákazníka, která je založena na transakční historii v průběhu sledovaného období, pomocí čehož určíme očekávaný počet budoucích transakcí pro náhodně vybraného zákazníka. Ten závisí na chování zákazníka v minulosti. Víme, že z různých důvodů může být zákazník vyřazen z našeho obchodního procesu: buď přestane nakupovat, odejde ke konkurenci nebo může vstoupit do likvidace. Ukážeme princip fungování modelu Pareto/NBD a popíšeme jeho matematickou i statistickou podstatu.

Ve třetí kapitole se zaměříme na praktickou aplikaci: popíšeme analyzovanou společnost, definujeme řešený problém, připravíme datový soubor a aplikujeme vybraný model Pareto/NBD na reálná data firmy „Práce pro Vás“ pomocí programovacího jazyka R a softwaru Rstudio. V této části práce je jedním z hlavních cílů vypočítat celoživotní hodnotu zákazníků firmy „Práce pro Vás“ pomocí Pareto/NBD modelu. Ukážeme, jak predikce CLV ovlivňuje

manažerské rozhodování a marketingovou strategii firmy. Na konci kapitoly je uvedena finanční analýza efektivity a zároveň je ukázáno, jak predikce CLV ovlivňuje manažerské rozhodování a marketingovou strategii firmy. Závěrečná kapitola představuje shrnutí výsledků. Demonstrujeme přínos, zhodnotíme využitelnost výsledků v praxi a uvedeme další náměty pro zkoumání.

2. Celoživotní hodnota zákazníka CLV

CLV je zkratka z angličtiny (the customer lifetime value), myslíme tím celoživotní hodnotu zákazníka pro společnost. Pro spoustu podniků se hlavní podnikatelský záměr pohybuje kolem snahy poznat klienta, kterého musí držet za každou cenu, a kterého ne. Dnešní společnosti se setkávají s rostoucí konkurencí a globalizací. Dnes je důležité být lepší a rychlejší než konkurence, je třeba se odlišit, rychleji reagovat na potřeby klienta či na podněty trhu a nacházet každý den nové řešení.

Existuje celá řada konceptů a přístupů, jak může firma odhalit zákazníka, který v nejbližším časovém horizontu odejde ke konkurenci a predikovat míru odchodu podobných zákazníků. To se pak dá využít pro rozhodující kroky k udržení klienta, nebo naopak zaměření na jiný klientský segment. Všechny modely a koncepty jsou doprovázeny výpočtem celoživotní zákaznické hodnoty, protože CLV poskytuje základ pro rozhodování o komunikační strategii a umožňuje zacílit marketingové nástroje na určitý klientský segment. Tuto problematiku hodně zkoumali pracovníci Univerzity Leuven (Glady, Baesens, Croux, 2009) a na základě daných teoretických podkladů popíšeme hlavní zásady konceptu CLV.

2.1 Popis konceptu CLV a příklad jeho výpočtu

Zkoumání a měření klientského chování je důležitou součástí pro každou komerční aktivitu. Každý podnik při strategickém plánování očekává určitý zisk a přínos ze své činnosti. Celoživotní hodnota zákazníka představuje ten samý budoucí výnos pro firmu. Za účelem predikce CLV potřebujeme zjistit očekávaný počet transakcí a sumu objednávek z těchto transakcí. Hodnota konkrétního klienta je velmi důležitá kvůli zjištění nejziskovějších vztahů pro společnost z komerčního hlediska. Takoví klienti vyžadují bližší kontrolu a sledování, na rozdíl od ostatních, kterým můžeme věnovat menší pozornost (z hlediska času i zdrojů). Tento přístup odpovídá principu racionálního využití zdrojů firmy. Z marketingového hlediska účinnější komunikační strategie předpokládá známou hodnotu zákazníka nebo cílové skupiny zákazníků pro kvalitnější aplikaci marketingových nástrojů. Téma oceňování zákazníka je popsáno Dwyrem a kol. (1987), kde hodnota zákazníka je založena na celkové době minulé spolupráce a minulém finančním stavu. Nicméně někteří autoři, jako například Kumar a Reinartz (2012), kritizovali tento přístup a demonstrovali na několika reálných příkladech, že CLV a minulé „partnerství“ občas nemají nic společného. Mezinárodní společnosti běžně využívají celoživotní hodnotu zákazníka jako nástroj pro měření úspěšnosti své činnosti. Podle šetření Yooa a Hanssense (2005) existuje několik příčin rostoucího zájmu o tento koncept.

Zaprvé je to zvyšování tlaku na marketingové oddělení s požadavkem na kvantitativní výstupy marketingové činnosti. Klasické marketingové ukazatele, jako povědomí o značce, vybavení, rozpoznání značky nebo objem prodeje, se jeví jako nedostatečné pro demonstraci návratnosti investic. Navíc autoři článku uvádí příklady fungování automobilového trhu, kdy aktivní marketingové úsilí má negativní účinek na ziskovost společnosti. Druhá příčina spočívá v omezené rozpoznávací schopnosti finančních ukazatelů, jako jsou cena akcií, celkový obrát nebo úhrnný zisk. Uvedené charakteristiky odráží pouze současný stav, neposkytují žádnou informaci o budoucím chování, které můžeme očekávat. Takže z praxe víme, že můžeme diferencovat zákazníky podle výnosu pro společnost, ale úhrnné finanční ukazatele nepředpokládají diferencovaný přístup. To je rozdíl oproti konceptu celoživotní hodnoty, kde hlavním krokem je přechod z celkových transakcí na individuální úroveň pro účely zjištění cílových klientů (Gupta, Lehmann a Stuart 2003). Zatřetí, inovace a technologický pokrok umožňují sbírat obrovská data o transakcích. Spousta firem zavádí věrnostní programy pro tento účel. Zákaznické karty jsou určeny především k zjišťování preferencí a sběru dat. Kromě toho v současné době společnost nepotřebuje náhodně definovat reprezentativní vzorek, je k dispozici celá zákaznická databáze, kterou za relativně krátký časový úsek může zpracovat dnešní software.

Problematikou se zabývali i Gupta a kol. (2006), kde autoři hodně času věnovali maximalizaci celoživotní hodnoty. Podle nich schématické modelování CLV vypadá následovně:



Obrázek 2.1 Koncept modelování CLV

Zdroj: Gupta a kol., Journal of Service Research (2006)

Na obrázku 2.1 můžeme vidět, že všechny marketingové aktivity ovlivňují zákazníka a směřují přes CLV k samotné hodnotě firmy. Daná schéma zdůrazňuje důležitost řízení a měření celoživotní hodnoty klienta, která je nástrojem oceňování současné komerční úspěšnosti a možnosti budoucího rozvoje byznysu. Už jsme zmiňovali, že existují odlišné pohledy na

odhadnutí celoživotní hodnoty. Například F. Reichheld, známý americký autor knih o strategickém marketingu a loajalitě zákazníka (2001), se hodně věnoval komponentním složkám firemní činnosti, které ovlivňují CLV nejvíce. Podle jeho teorie je udržení zákazníka tím nejdůležitějším záměrem firmy. Ale neexistují zde přímé lineární vztahy mezi těmito ukazateli: klienti, se kterými pracujeme nejdelší dobu, nemusí být nejziskovějšími. Takže Gupta a kol. zaměřili pozornost na 5 finálních složek firemní hodnoty. Z toho lze usoudit, že celoživotní hodnota zákazníka může plnou měrou zastupovat odhad hodnoty firmy. Ve svém šetření Kumar a kol. (2008) demonstrují silnou korelační závislost mezi CLV a oceněním společnosti prostřednictvím longitudinální analýzy aktivit podniku. Termín „longitudinální“ vysvětluje Michael Berbaum (2012): dlouhotrvající analýzy nebo opakované šetření, což v širokém slova smyslu je obdobou panelových výzkumů, kde jednu skupinu subjektů opakovaně zkoumáme v delším časovém horizontu.

Spousta autorů se zabývá i dalšími složkami měření celoživotní hodnoty zákazníka. Například výše zmínění Srinivasan a Hanssens (2007) popisují spojení nákladů na reklamu a cenu akcií společnosti bez ohledu na objem prodeje a počet transakcí. Existují také šetření a business analýzy, které zkoumají závislost CLV na subjektivních faktorech, jako jsou například spokojení zákazníci a síla značky (Kumar, Luo a Man 2006). Tyto faktory bezpochyby pozitivně ovlivňují dobu trvání vztahů s klientem. Ale, jak už jsme uvedli dříve, délka vztahu nemusí být jediným ukazatelem hodnoty zákazníka, protože subjektivita předpokládá pouze úmysl zakoupit než skutečný nákup.

Dále budeme pokračovat v našem výkladu a uvedeme fundamentální základy modelování CLV a ukážeme několik způsobů odhadu této hodnoty.

Celoživotní hodnotu zákazníka různí autoři definují odlišně, ale potřebám této práce zcela odpovídá definice, kterou uvedl Kumar (2008). CLV je současná hodnota všech budoucích příjmů generovaných zákazníkem. Celoživotní hodnota je také podobná diskontovanému peněžnímu toku podle finančního přístupu. Nicméně existují dva hlavní rozdíly mezi nimi. Zaprvé, CLV je vymezená a odhadovaná na individuální úrovni zákazníka nebo jednotlivého segmentu. Tento přístup umožňuje diferencovat marketingová úsilí a efektivně alokovat zdroje ve srovnání s měřením průměrných ukazatelů pro celý datový soubor. Zadruhé, na rozdíl od finančního hlediska, CLV přímo zahrnuje pravděpodobnost přechodu ke konkurenční firmě v budoucnu.

CLV můžeme spočítat podle následujícího vzorce (Reinartz, Kumar 2004):

$$CLV = \sum_{t=0}^T \frac{(p_t - c_t)r_t}{(1+i)^t} - AC, \quad (2.1)$$

kde

p_t – cena objednávky v okamžik t ;

c_t – náklady na danou objednávku v čase t ;

i – úroková míra;

r_t – pravděpodobnost opakované objednávky nebo pravděpodobnost, že zákazník je aktivní v čase t ;

AC – náklady na akvizici zákazníka;

T – časový horizont odhadu CLV.

Navzdory relativně jednoduchému vzorci dochází často k problémům při zjišťování nákladů na akvizici zákazníka nebo stanovení správného časového horizontu. Podle Gupty a Lehmannna (2010) platí, že máme-li k dispozici náhodný nebo dokonce i neomezený časový horizont, tak získáme nadhodnocenou hodnotu CLV (občas docela značně).

Výše uvedení autoři také ukazují, že pokud platí následující podmínky - hodnota marže $p_t - c_t$ (rozdíl mezi prodejní cenou a náklady) je konstantní v čase t ; můžeme využít neomezený časový horizont a hodnota retence r_t je také konstantní v čase t – tak lze vzorec pro výpočet CLV zjednodušit do následujícího tvaru:

$$CLV = \sum_{t=0}^{\infty} \frac{(p-c)r^t}{(1+i)^t} = m \frac{r}{1+i-r}, \quad (2.2)$$

kde m je hodnota marže. Jinými slovy, CLV je m -násobek hodnoty $r/(1+i-r)$. V tomto okamžiku musíme zdůraznit, že většina modelovacích přístupů ignoruje konkurenční prostředí a rychle měnící se okolí podniku. Frekvence aktualizace celoživotní hodnoty záleží na dynamice určitého trhu. Jednoznačně jsou vidět rozdíly mezi trhem mobilních telefonů nebo mléčných výrobků. A proto manažeři určují marketingovou strategii a přístupy k odhadu CLV dle svých vlastních potřeb. Ze stejného důvodu existuje množství přístupů k složkám ovlivňujícím CLV mezi jednotlivými autory. Gupta a kol. (2006) nabízejí šest různých přístupů k modelování celoživotní hodnoty. V následující subkapitole popíšeme některé z nich podrobněji.

2.2 Způsoby modelování CLV

V této subkapitole krátce charakterizujeme šest různých modelů a přístupů, které se používají k odhadu CLV, a vybereme konkrétní model pro aplikaci na reálných datech společnosti. I když každý model řeší v zásadě stejnou otázku (který zákazník je nejziskovější

a jak alokovat zdroje podniku nejúčinnějším způsobem pro dosažení maximálního zisku), zmíníme přesto rozdíly mezi následujícími přístupy.

RFM modely

RFM je zkratka v anglickém jazyce, která znamená *recency*, *frequency*, *monetary*. Pod pojmem *recency* máme na mysli lhůtu poslední aktivity klienta. Čím menší je ta lhůta, tím je větší pravděpodobnost opakování objednávky. Další hodnotou modelu je frekvence, počet uskutečněných transakcí. A nakonec je tu peněžní ukazatel – *monetary* si lze představit jako sumu objednávek za sledované období. RFM modely popisují Fader, Hardie a Lee (2005). Daný typ modelování se používá v přímém (direct) marketingu posledních 30 let především pro přesné zacílení na užší nebo dokonce i na jednotlivé segmenty. Prioritním záměrem je zvýšit tzv. response rate. Tento ukazatel je charakterizován v knize od Karlíčka a Krále (2011). „Jedná se o počet příjemců sdělení, kteří na nabídku zareagovali v relaci k celkovému počtu oslovených“. Často společnosti využívají demografické charakteristiky pro cílení na zákazníka, ale Fader a kol. (2005) moc nedoporučují nahrazovat minulé chování za demografický profil pro predikce budoucího chování. RFM modely fungují následujícím způsobem: nejprve roztrídíme zákazníky na skupiny podle třech proměnných – *recency*, *frequency* a *monetary*. Nejjednodušší modely zařazují klienti do pěti skupin, založených na těchto třech proměnných. Získáváme $5^3 = 125$ buněk, kterým přiřadíme jednotlivé váhy. Každá skupina má určité ohodnocení, podle kterého alokujeme zdroje a upravujeme nástroje přímého marketingu. Pomocí tohoto principu lze predikovat budoucí chování a pomocí ohodnocení model směřuje k CLV, avšak jsou zde některá omezení (Kumar, 2008). Zaprvé RFM modely predikují hodnoty pouze na jedno období. Nicméně společnosti odhadují CLV pro celý časový úsek (0, T]. Druhé omezení spočívá v nedokonalé predikci skrytého, podmíněného chování, protože model je založen na prostém třídění do skupin. Dále také RFM modely neberou v potaz skutečnost, že aktivita zákazníka může být způsobena minulými marketingovými nástroji. I přes tato omezení dané modely zůstávají populárním nástrojem pro predikce hodnoty zákazníka, především díky jednoduché realizaci, která nevyžaduje ani sofistikovaný software a ani speciální technické znalosti. Samotný proces seskupení zákazníků a hloubková analýza minulého chování je kvalitním způsobem, jak predikovat nákupní chování v budoucnu. Fader, Hardie a Lee (2005) popisují, jak je možno použít R, F, M skupiny a zároveň překonat uvedená omezení. Důležité je to, že RFM modely poskytují postačující statistiky pro predikci CLV.

Na základě šetření (Jašek a Vraná, 2014) sledujeme rozdíly mezi modely: „Každý model slouží k určitému účelu a může být aplikován na různé firmy podle rozmanitých podnikatelských záměrů. RFM modely jasně vizuálně demonstrují dopad jednotlivých složek, když pomoci pravděpodobnostních modelů získáváme pravděpodobnost aktivity zákazníka, vektorové autoregresní modely naznačují možné dynamické vztahy mezi proměnnými”. Dále článek uvádí informace o tom, že RFM modely ukazují důležitost prvních dvou složek, které jsou klíčové pro predikci CLV. Poslední složka – suma objednávek – vůbec nemusí určovat budoucí transakci. Pokud se na to podíváme z jiného úhlu pohledu, tak vysoká pravděpodobnost nákupu se objevuje u klientů, které nakupovali velmi často v poslední době a na základě těchto informací manažeři společnosti rozhodují o zvýšení intenzity marketingového úsilí.

Ekonometrické modely

Další třída modelů je spojena s pravděpodobnostním modelováním (analýza přežití) a odhadem nákupního chování. V podstatě dané modely kombinují tři hlavní složky z obrázku 2.1 (akvizice, udržení a navýšení prodeje) pro odhad CLV.

Akvizice (pořízení) nového klienta se vztahuje k prvnímu nákupu. Šetření v této oblasti je zaměřeno na analýzu rozhodování o nákupu a pokusech vysvětlit spojení mezi udržením zákazníka a jeho prvním nákupem. Thomas, Kumar, Reinartz (2005) tvrdí, že rozhodnutí o akvizici nebo neakvizici určitého zákazníka musí být založeno především na odhadu budoucí ziskovosti, ne na odhadu nákladů na akvizici. Lewis (2004) ukazuje na příkladu novin, jak propagace, která zvyšuje počet nových zákazníků, může z dlouhodobého hlediska škodit. Jestliže klient novin, který si předplatil tisk, bude ochoten nabízet běžnou cenu, pravděpodobnost opětovného zakoupení bude kolem 70 %. Nicméně pravděpodobnost se prudce zmenšuje na 35 %, pokud bude zákazník získán pomocí týdenní slevy na noviny.

Rozlišujeme dva základní typy modelování udržení klienta. První typ považuje přechod zákazníka ke konkurenční firmě za trvalý jev a snaží se odhadnout pravděpodobnost toho přechodu. Druhý typ považuje změnu firmy dodavatele služby nebo výrobku za tranzientní stav, do kterého se už nikdy nevrátí. Tento proces je popsán pomocí Markovských řetězců (Kořenář, 2002). Někteří autoři (Rust, 2004) popírají takový přístup, protože návrat zákazníka je úplně ignorován, ale klienta můžeme často znovu přilákat. Vrátime-li se k nástrojům direct marketingu, tak Kumar (2008) zjistil, že četnost kontaktování jednotlivých zákazníků má pozitivní, ale nelineární vliv na četnost nákupu. Reinartz, Thomas a Kumar (2005) píší, že

osobní komunikace pozitivně ovlivňuje celoživotní hodnotu, spotřebitelské výdaje a loajalitu klienta. Třetí složkou CLV je navýšení prodejů a ziskové rozpětí neboli marže (rozdíl mezi náklady a nákupní cenou), generovaná spotřebitelem za časový úsek t . Ziskové rozpětí závisí jak na minulém chování, tak i na marketingovém úsilí firmy. Pro zachycení změn v marži s časem Venkatesan i Kumar (2004) využili jednoduchý model lineární regrese:

$$\Delta_{jt} = \beta x_{jt} + e_{jt}, \quad (2.3)$$

kde změna marže pro j -tého zákazníka v čase t je vyjádřena funkcí (2.3) a e_{jt} představuje odchylku odhadu náhodné složky, β je neznámý parametr regresní funkce.

Modely založené na analýze časových řad

Podobně jako ekonometrické modely se modely založené na analýze časových řad zaměřují na dynamiku určitých komponent (akvizice, udržení apod.). Za určitých podmínek, jako jsou například dostatečně dlouhodobé sledování nákupního chování a dostatek informací o klientech, můžeme zacházet s jednotlivými složkami jako s jediným celkem a dynamickým modelem. Tento přístup je také široce využíván i v kontextu celoživotní hodnoty pro analýzu vlivů reklamy, slevové politiky, kvality výrobku i služeb na zákaznickou hodnotu.

Základem je aplikace vektorového autoregresního modelu (VAR) pro predikci zákaznické hodnoty (Villanueva a kol., 2008). Autoři zkoumali vliv akvizice klientů na celkový výkon společnosti pomocí odlišností mezi zákazníky pořízené marketingovými aktivitami a náhodně získanými zákazníky. Model je založen jako klasický VAR (p) model, kde p je počet zpoždění a zachycuje dynamické vztahy mezi třemi proměnnými: počtem klientů získaných pomocí marketingových úsilí (MKT), počtem zákazníků získaných pomocí ústního sdělení (WOM) a výkonem firmy – VALUE (Jašek, Vraná, 2014). Podle provedeného šetření (Villanueva a kol., 2008) zákazníci získaní pomocí reklamy a propagace měli vyšší hodnoty z krátkodobého hlediska, ale náhodně pořízení zákazníci výrazněji ovlivnili výkon firmy v dlouhodobém horizontu. Tuto problematiku na základě daného přístupu (Villanueva a kol., 2008) hodně zkoumali výše uvedení autoři (Jašek a Vraná, 2014) a dospěli k velmi zajímavým výsledkům. Na reálných datech získaných od maloobchodu módních oděvů prokázali, že výsledky modelování se mohou lišit a klienti pořízení na základě WOM komunikace nemají až tak významný dopad na výkon celé společnosti. Podle daného výzkumu klienti, získaní pomocí ústního sdělení, obvykle nakupovali pouze jednou, což je naopak ve srovnání se závěry výzkumu Villanuevy a kol. Z tohoto výzkumu vyplývá, že WOM je hlavní složkou, která ovlivňuje firemní výkon a hodnotu.

Různé přístupy k modelování CLV umožňují získat celkový přehled a porovnat výsledky různých algoritmů mezi sebou (Yoo a Hanssens, 2005). Hlavní přínos modelování, založeného na analýze časových řad, spočívá v dlouhodobém a rovnovážném chování proměnných nebo skupin proměnných. Například můžeme sestavit model dohromady z několika populárních marketingových nástrojů. Pokud společnost provede marketingovou kampaň a získá nové zákazníky, tak tento úspěch poslouží jako podnět pro investici do dalších marketingových pomůcek. Finanční prostředky na novou kampaň může společnost přesměrovat z jiných částí marketingového mixu, což hodně ovlivňuje ostatní kanály a vnitřní okolí vcelku. Najednou vnější okolí firmy (konkurenti, distributoři, dodavatelé) také reaguje na změny, například na pokles tržního podílu u konkurentů a začíná vyvíjet propagační a reklamní aktivitu, čímž způsobuje změny na celém trhu. Záleží na tom, jak silně reaguje na okolí podniku a na samotnou velikost podniku. V rozsáhlé změně jak vnitřního, tak i vnějšího prostředí firmy spočívá systémová dynamika, která nám říká, že všechny složky jsou navzájem propojené. Navíc tyto systémy umožňují kvantifikovat relativní důležitost (váhu jednotlivých složek) systému v dlouhodobém horizontu. Ale musíme zmínit, že výsledky modelování silně korelují s kvalitou dat, která máme na začátku. Jak je uvedeno v navazujícím článku (Jašek a Vraná, 2015), zaměřeném na modely založené na analýze časových řad, jednou z hlavních výhod je možnost predikovat hodnoty všech tří komponent (MKT, WOM, VALUE). Autoři znovu demonstrují na několika reálných firmách, že stejný aplikovaný přístup může vést k odlišným výsledkům modelování: očekávaný růst počtu nákupů byl způsoben převážně MKT zákazníky, nicméně účinek ústního sdělení nebyl zatím prokázán. Ve skutečnosti na příkladu kosmetického maloobchodu bylo prokázáno, že akvizice nových zákazníků takovým způsobem negativně ovlivnila počet objednávek.

Daný typ modelování umožňuje kvantifikovat jak sílu, tak i délku marketingového vlivu na výkon firmy, což je velice důležitý výstup pro budoucí rozhodování (Dekimpe, Hanssens, 2005).

Machine learning

Machine learning (též strojové učení) znamená, že systém se učí samostatně a zlepšuje výstupy modelování s každým krokem podle stanoveného algoritmu. Jak uvádí někteří autoři (Jangid a Kathari, 2014), hlavním přínosem strojového učení je zaměření na CE (Customer Equity – celková hodnota CLV pro všechny zákazníky), což umožňuje získat komplexní přehled o transakční aktivitě a klientském chování. V podstatě CE je suma celoživotní hodnoty

skutečných a budoucích zákazníků. Marketéři často preferují jasně vymezené parametrické modely (zmíněné výše). Takové modely jsou založeny na teorii a můžeme je relativně lehce přizpůsobit a interpretovat. Což je rozdíl oproti počítačovým neparametrickým statistikám, které generují předpovědní schopnosti, například projekce nebo počítačové programy napodobující lidské myšlení či neuronovou síť, stromové modely, které pocházejí z data miningových technik (Friedman, 1991), apod. Příklad docela úspěšného použití data miningových technik sledujeme v šetření (Soeini a Fathalizade, 2012), kde autoři navrhli zcela nový princip segmentace klientů a zacílení marketingových nástrojů na internetový obchod hraček pomocí lineární regrese. Na základě výpočtu CLV autoři rozdělili klienty na tři základní segmenty a pro každý segment určili jednotlivou cenovou politiku. Nakonec dospěli k závěru, že obchod potřebuje snížit ceny a rozšířit nabídku služeb pro stálé zákazníky, které preferují nižší ceny, aby se zvýšil zisk firmy. Takovým způsobem manažeři společnosti stanovili, že klienti v segmentu s nejnižší cenou mají největší celoživotní hodnotu a přináší společnosti mnohem větší zisk, než klienti, kteří kupují luxusní výrobky za vysokou cenu, ale pouze jednou bez opakování. Daný příklad ukazuje důležitost modelování CLV a dopady na marketingovou strategii firmy.

Co se týká hlavních počítačových programů, kterých se zde využívá, je potřeba zmínit GAM (generalized additive model), MARS (multivariate adaptive regression splines), CART (classification and regression trees) a SVM (support vector machines). Většina těchto přístupů je vhodná pro analýzu mnohem většího počtu proměnných, kdy variabilita dat je tak velká, že se odhad parametrů jeví jako nemožný a nemůžeme spoléhat na parametrické a ani neparametrické testy. K překonání podobných situací slouží GAM modely, které uvedli poprvé Hastie a Tibshirani (1990). Hlavní myšlenka modelu spočívá ve skutečnosti, že střední hodnota závislé proměnné závisí na takzvaném aditivním indikátoru, který je stanoven pomocí nelineární neparametrické funkce. Trochu jiný přístup nabízí model MARS – využívá neparametrického regresního algoritmu, který funguje na principu vícenásobné lineární regrese s kontrolními body, body odhadujeme z dat pomocí vhodného softwaru (Friedman, 1991). SVM je později vzniklý model, který byl zaveden především pro zjednodušení klasifikace a třídění dat. Namísto předpokladu o lineární regresi, která rozděluje dvě nebo více tříd, tento přístup umožňuje vysvětlit nelineární vztahy. SVM modely efektivně transformují primární data do upraveného tvaru pomocí matematických a statistických výpočtů tak, že můžeme s nimi pracovat pomocí lineárních metod (Friedman 1991). Každá společnost se musí rozhodnout podle svých vlastních potřeb, který model zvolit, ale vždy má možnost zkvalitnit výsledky predikce pomocí sloučení několika přístupů. Počítačové modely fungují účinně,

protože zlepšují přesnost jediného modelu, při tom používají mnohonásobné kopie modelu na různé části dat. V literatuře velmi často potkáváme termín machine learning nebo strojové učení. „Je to součást umělé inteligence, kdy se počítačový systém pomocí sofistikovaného algoritmu učí samostatně, pod učením rozumíme vnitřní změnu systému, která zefektivní schopnost přizpůsobení se změnám okolního prostředí. Strojové učení má široké uplatnění v oblasti statistiky, matematiky, podpory rozhodování. Hlavní typy úloh jsou klasifikace vstupních dat do tříd, odhad číselné hodnoty výstupu podle vstupu, shlukování (Witten, Frank, Hall, 2011).

Modely růstu

Celoživotní hodnota je dlouhodobá očekávaná ziskovost jednotlivých zákazníků. Jak už víme, CLV se používá v marketingu a managementu pro segmentaci, zacílení a alokaci zdrojů (Kumar, 2004). Ze strategického hlediska musí být analýza CLV funkční a poskytovat metriky i pro top manažery společnosti. S ohledem na uvedenou myšlenku několik výzkumníků navrhuje zaměřit se na CE (customer equity), což je celoživotní hodnota všech aktuálních a budoucích zákazníků (Gupta a Lehmann, 2010). Predikovat hodnotu nově získaných klientů můžeme dvěma způsoby. První přístup postupně rozděluje data do menších tříd a pak poskytuje ukazatel pravděpodobnosti získání konkrétního zákazníka. Alternativní postup naopak naše data agreguje do jediného celku, kterému říkáme „narůstající model“ a predikuje počet zákazníků, které firma může získat v budoucnu. Někteří autoři používají daný typ modelování i pro ohodnocení ztracených zákazníků (Hogan, Lemon, Libai, 2003). Výzkumníci popisují, že ztráta klienta znamená nejenom pokles zisku, ale i negativní WOM (word of mouth – ústní sdělení) efekt, což nese s sebou další dlouhotrvající následky. Například v oblasti internetového bankovníctví odhad přímé ztráty činil 208 dolarů, zatímco vedlejší dopad dosáhl hodnoty 850 dolarů. Hlavním přínosem modelu růstu je možnost kvantifikovat nově získané zákazníky, a tudíž i budoucí příjmy společnosti (EsmaeiliGookeh a Tarokh, 2013).

Pravděpodobnostní modely

Nakonec jsme se dostali k šestému typu modelování CLV, který bude v další části práce středem našeho zájmu. Pravděpodobnostní modely jsou znázorněním pozorovaných procesů, které jsou výsledkem realizace skrytých stochastických jevů, řídicích nahodilostmi a různých od zákazníka k zákazníkovi. Středem pozornosti v procesu budování těchto modelů je popis pozorovaného chování místo pokusů vysvětlit rozdíly a zákonitosti v chování, jak je to u

regresních modelů. Výzkumníci v daném oboru předpokládají, že spotřebitelské chování se liší u různých populací podle některých pravděpodobnostních rozdělení (Kumar a Shah, 2015). Naším cílem v rámci CLV je predikovat, bude-li náš zákazník aktivní v budoucnu, a pokud ano, jaké nákupní chování od něj můžeme očekávat v příštím období. Existuje řada pravděpodobnostních modelů, ale první pravděpodobnostní model, který přímo řešil danou problematiku, je model Pareto/NBD, kterému budeme věnovat samostatnou subkapitolu. Zkratka NBD znamená negativní binomické rozdělení pravděpodobnosti. První šetření v této oblasti byla realizována s cílem popsat transakční tok a stanovit kvantitativní charakteristiky počtu a frekvence objednávek pro další zpracování a predikce (Schmittlein, Morrison, Colombo, 1987). Díky prudkému rozšíření pravděpodobnostních modelů výzkumníci rozpracovali i další alternativy modelování náhodných procesů: model BG/NBD (Fader a kol. 2005) a model BG/BB (Fader a kol., 2010). Ačkoliv základna uvedených modelů je stejná, tak jsou mezi nimi rozdíly především v interpretaci neaktivní doby chování klienta. Paretův model předpokládá odchod zákazníka v kterýkoliv okamžik z hlediska času, bez ohledu na vyskytující se objednávky v současném období. Ale pokud budeme předpokládat, že klient odejde okamžitě po nákupu, můžeme popsat daný náhodný proces pomocí beta geometrického (BG) modelu. Po každém nákupu se stane zákazník neaktivní s pravděpodobností p . Z tohoto důvodu časové okamžiky odchodu jsou znázorněny geometrickým rozdělením. Heterogenita tohoto jevu je popsána beta rozdělením pravděpodobnosti. Podle doporučení autorů modelu (Fader, Hardie, Lee, 2005) každý podnik musí začít modelování přístupem BG/NBD v případě, že přesně neví, jak se rozhodnout. „Model BG/NBD je správný počáteční bod pro jakoukoliv zákaznickou analýzu, kde transakce jsou spojitě a okamžik odchodu klienta je stochastický proces“.

Co se týká modelu BG/BB, název pochází z kombinace beta geometrického a beta binomického rozdělení. Logika věci je shodná s ostatními modely, ale časový okamžik výskytu transakcí klienta, jak je zřejmé z binomické části pravděpodobnostního rozdělení, je diskrétní náhodná veličina, tj. intervaly mezi uskutečněnými objednávkami jsou fixní a pravidelné (Fader a kol., 2005). Model BG/BB předpokládá heterogenitu jak pro pravděpodobnost nákupu, tak i pro začátek neaktivního období.

Současný trend v marketingu je snaha kvantifikovat výstupy. Pomocí vhodných metrik a modelů má marketingové oddělení možnost vyhodnotit svou práci a ukázat návratnost investic do komunikačních prostředků. CLV považujeme za jednu z takových nejvýznamnějších metrik, která je relativně snadno pochopitelná v teorii a aplikovatelná v praxi.

Popsané modelování CLV a sofistikovaný software nakonec umožňují marketérům získat kvantitativní výstupy, u kterých společnost upravuje marketingovou strategii podle potřeb konkrétních zákazníků. Faktem je, že každá společnost musí udělat ten správný výběr, závisející na oboru podnikání (banky, výroba, služby, strojírenství apod.). Například existují monogamní společnosti, spolupracující pouze s jediným zákazníkem. Ale v neméně častých případech společnost současně komunikuje se spoustou různých zákazníků a vždy se řeší otázka výběru cílového klienta.

Výše uvedené modely mají různé vstupy, podmínky, omezení a samozřejmě odlišné výsledky modelování. V článku (Jašek a Vraná, 2014) autoři podrobně popisovali tři typy modelů – RFM, pravděpodobnostní modely i modely založené na analýze časových řad - a porovnávali výstupy modelování mezi sebou. Z této práce můžeme odvodit, že porovnání modelů v praxi je značně obtížné: RFM modely využívají zákaznická data a predikují CLV na individuální úrovni, naproti tomu při analýze časových řad zkoumáme agregovaná data. Ačkoliv jsou cíle u všech modelů stejné, tak nakonec autoři šetření získali odlišné výsledné hodnoty z výše uvedených typů modelování.

Existují i jiné přístupy a algoritmy výpočtu CLV. Například v článku (Flordal, Friberg, 2013) je uvedeno pět typů modelování: prosté deterministické modely, ve kterých sečteme dodatečný cash flow (peněžní tok); modely založené na udržení zákazníka, kde je cílem určit pravděpodobnost odchodu zákazníka v čas t ; takzvané „migrační“ modely, kde je odmítán předpoklad o tom, že pokud zákazník neuskutečnil nákup v průběhu určitého času, tak odešel jednou provždy. Podle předpokladu o „migraci“ klienta mezi konkurenčními firmami stanovíme pravděpodobnost budoucího nákupu na základě informace o poslední transakci. Ačkoliv i doba poslední objednávky ovlivňuje skutečný stav: čím déle klient nekupuje výrobky nebo služby společnosti, tím více klesá pravděpodobnost příštího nákupu (Nikkhahan, Hadrabadi, Tarokh, 2011). Dále autoři (Flordal, Friberg, 2013) uvádějí dva typy modelování založené na Markovských řetězcích. Na základě Markovských řetězců byl vypracován model, který umožňuje stanovit pravděpodobnost přechodu jednotlivých klientů do konkurenční firmy (Rust a kol., 2004) pomocí takzvané fundamentální matice přechodu mezi jednotlivými stavy – ty jsou buď absorpční nebo tranzientní (Kumar a George, 2007).

Jak je vidět z této podkapitoly, existuje spousta metod a přístupů k dané problematice. Ale každý model prokazuje nutnost analyzovat své zákazníky a přizpůsobovat marketingové nástroje jednotlivým segmentům. Pro efektivní alokaci zdrojů marketéři rozhodně nesmí zapomenout na predikce CLV a vybrat nejvhodnější způsob modelování. Úplný přehled a popis všech populárních přístupů v této oblasti nabízejí tito autoři (EsmaeiliGookeh a Tarokh, 2013).

Článek popisuje významné šetření v oblasti predikce zákaznické hodnoty od roku 1994. Pro lepší přehlednost je informace zobrazena ve formě tabulky a články jsou seřazeny chronologicky. Následující tabulka obsahuje i hlavní přínos jednotlivých šetření v třetím sloupci.

Tabulka 2.1 Přehled šetření v oblasti modelování CLV

Autoři	Rok vydání	Klíčový přínos
Wang, P. and Spiegel, T.	1994	Dynamické působení mezi náklady na akvizici a udržení klienta
Carpenter	1995	Kvantifikace vztahů mezi společností a zákazníky
Blattberg and Deighton	1996	Určení nákladů na akvizici zákazníka
Berger and Nasr	1998	Výpočet čisté marže
Colombo and Jiang	1999	Stochastické RFM modely
Pfeifer and Carraway	2000	Markovské řetězce
Rust	2000	Markovské řetězce a odchod ke konkurenci
Blattberg	2001	Návratnost investic na akvizici, udržení a add-selling
Hwang, Jung, Suh	2003	Skutečná hodnota, budoucí hodnota a loajalita
Liu, Zhao, Zhang and Lu	2004	Word-of-mouth
Venkatesan and Kumar	2004	Alokace zdrojů
Fader, Hardie, Lee	2005	RFM
Hughes	2005	RFM
Hwang, Jung, Suh, Kim	2006	Křížový prodej a ovlivňování CLV
Crowder, Hand and Krzanowski	2007	Očekávaný výnos z marketingových aktivit
Haenlein, Kaplan and Beaser	2007	Markovské řetězce a CLV
I-Cheng Yeh, King-Jang Yang, Tao-Ming Ting	2009	Rozšíření RFM na RFMTC a LRFMC model
Cheng, Chiu and Wu	2011	Markovské řetězce a CLV
Ahmadi, Taherdoost, Fakhravar and Jalaliyoon	2011	Riziko a možnost přizpůsobit se; náklady na akvizici a udržení zákazníka

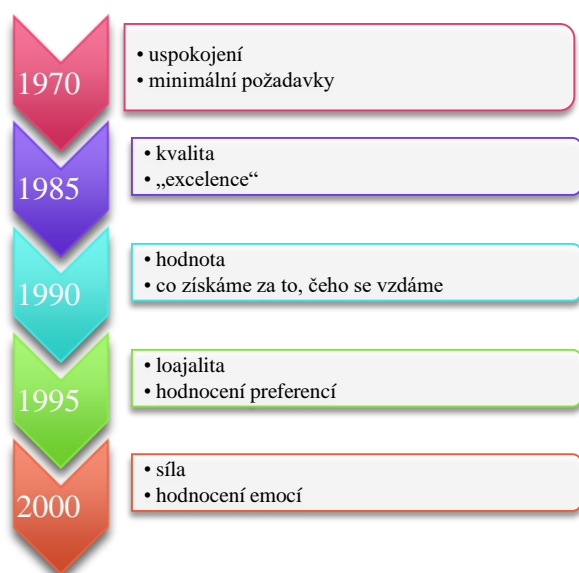
Zdroj: EsmaeiliGookeh, Tarokh: Customer Lifetime Value Models: A literature survey (2013)

2.3 Koncept řízení vztahů se zákazníky CRM

Během práce především mluvíme o CLV, ale v tomto případě nemůžeme nezmínit pojem CRM. Termín CRM je anglická zkratka – Customer Relationship Management – která se dá přeložit jako řízení vztahů se zákazníky. Tento koncept je aktuální a zásadní nástroj pro daný účel. Je třeba pečovat o významné zákazníky více než o zákazníky okrajové, jejichž nákupy jsou náhodné bez určení dlouhodobých vzájemných kontaktů. Kvůli výraznému vývoji informačního prostředí, rozvoje konkurence na trzích a rychlých změn v podnikatelském prostředí nemá žádná firma záruku neměnné struktury svých zákazníků. Z tohoto faktu vyplývá, jak je důležité se v současné době věnovat zákazníkům, diferencovat a segmentovat přístupy k nim a pravidelně vyhodnocovat, kontrolovat a korigovat firemní chování a péči o zákazníka. Tradiční přístupy k marketingu a komunikacím se neustále mění. Manažeři vyžadují stále nové metody, modely a druhy šetření v této oblasti. Díky tomu je strategie podniku nastavená tak, aby byli osloveni a udrženi důležití zákazníci. Zákazníci jsou ústředním bodem, kde se propojí všechny zdroje, které má podnik k dispozici. Pro tvorbu dlouhodobých vztahů, poskytování kvalitního servisu a udržování podílu na trhu je zcela zásadní implementace komplexní CRM strategie pro řízení vztahů se zákazníky. Daným konceptem se hodně zabývají autoři odborné literatury o tomto tématu (Reinartz, Krafft, Hoyer, 2004). „Zavádění CRM programu bez konceptu CLV je pro společnost obdobné nacházení uprostřed cesty s nejasnou představou, kam se posunout dál“ (Journal of Marketing Research, 2004). Citovaná věta problém velmi dobře popisuje a naznačuje vztah mezi celoživotní hodnotou a základním procesem CRM. Komplexní přístup umožňuje firmě transformovat CRM pomocí CLV v kvalitní a výkonný proces řízení vztahů se zákazníky a získat nepopíratelnou konkurenční výhodu.

V literatuře se dají najít různorodé definice daného procesu. CRM je integrovaný podnikový proces pro diferencovanou komunikaci s jednotlivými segmenty zákazníků (Baran a Galka, 2013). Jsou to programy, které umožňují shromažďovat, třídit a zpracovávat údaje o zákaznících, především jejich kontakty, probíhající obchodní procesy a dosahované tržby. CRM systémy tak pomáhají sledovat a vyhodnocovat veškeré obchodní aktivity v rámci celé společnosti. I z tohoto důvodu bývají součástí CRM systémů nejrůznější statistiky. CRM se stává klíčovou záležitostí pro firmy a organizace všech velikostí. V období devadesátých let se stala dalším objektem měření kvalita a její vnímání ze strany zákazníka. Po tomto kroku se pozornost přesunula k loajalitě zákazníka. Pokud má firma základnu loajálních zákazníků, kteří důvěřují značce, mají značku rádi a pravidelně od ní kupují výrobky nebo služby, tak je to

zřejmě pevná konkurenční výhoda. Loajální klienti jsou aktivem, jehož hodnotu se snažíme spočítat, analyzovat a predikovat. Ale musíme si všimnout, že loajální zákazníci nejsou samozřejmostí, vždy bereme v potaz plán, jak je udržet a jakou máme konkurenční výhodu pro budování dlouhodobých kvalitních vztahů (Kotler, 2000). Obrázek 2.2 byl vytvořen na základě teoretických podkladů (Lehtinen, 2007) a schematicky popisuje vývoj CRM podle výzkumu zákazníků.



Obrázek 2.2 Vývoj CRM systému

Zdroj: Aktivní CRM. Řízení vztahů se zákazníky, 2007

Na výše uvedeném obrázku sledujeme postupný přechod od výrobního pojetí, kde se strategie pohybovala kolem prodeje, k zákaznický orientované společnosti, kde hlavním cílem je spokojenost klienta. Ovšem musím zmínit, že měření spokojenosti zákazníka je dost složitá záležitost. Je totiž přirozeně těžké oddělit ji od celkových výsledků podnikání. Zejména je obtížné najít příčinné souvislosti. V současné době je zákazník považován za subjekt vztahu. Marketéři se snaží budovat model partnerských postojů, kde zákazník určuje marketingovou strategii. Za hlavní výsledky této subkapitoly lze považovat myšlenku, že úspěšnost nebo neúspěšnost CRM mechanismu celého podniku posuzujeme na základě změny celoživotní hodnoty zákazníka. Popsali jsme různé přístupy a klíčové pojmy spojené s celoživotní hodnotou zákazníka, což nám umožňuje vybrat konkrétní model, který odpovídá potřebám této bakalářské práce a bude použitelný na reálná data, které máme k dispozici. Následující subkapitola bude věnovaná pravděpodobnostnímu modelu Pareto/NBD. S jeho pomocí se dá predikovat budoucí počet objednávek a vypočíst CLV, kterou je často definována jako

současná hodnota očekávaného počtu transakcí přepočítaného na budoucí zisk z jedné transakce (Glady, Baesens, Croux, 2004).

Na závěr této subkapitoly zmíníme, že při správném aplikování jednotlivých modelů manažeři společnosti získávají užitečný přehled dopadů marketingových nástrojů a vnitřního mechanismu fungování podniku, což nese neocenitelný užitek pro proces manažerského rozhodování (Jašek a Vraná, 2015). Marketingová strategie se zaměřuje na rozvoj trvalé konkurenční výhody pro firmu nebo značku. Z tohoto důvodu je důležitým aspektem marketingových šetření především soustředění na dlouhodobý dopad všech marketingových nástrojů na firemní hodnotu. Jak už bylo zmíněno, modely založené na analýze časových řad se věnují důležitému principu, že komerční a marketingový úspěch záleží na integraci zákaznického chování, konkurenčního prostředí a na samotné společnosti. Firma musí pečlivě zhodnocovat a analyzovat celou řetězovou reakci s časem za účelem kvantifikovat své marketingové aktivity a zpracovat komplexní efektivní strategii. Proto musí marketéři určit vhodný model, který odpovídá potřebám konkrétní společnosti (Dekimpe, Hanssens, 2005). Longitudinální analýzy stále prokazují svou kvalitu a predikční přesnost, což zajišťuje široké použití a rostoucí popularitu pro různá odvětví a trhy.

2.4 Popis modelu Pareto/NBD

S ohledem na teoretické předpoklady, uvedené v minulé podkapitole, zásadní problém pro dnešní manažery je predikovat budoucí chování klienta, získat kvantitativní podstatu, jinými slovy – vstupní hodnoty pro výpočet celoživotní hodnoty zákazníka. V této části práce se budeme zabývat popisem a aplikací modelu, který se nazývá Pareto/NBD. Popis, matematickou i statistickou podstatu a vysvětlení fungování modelu uvádí Peter S. Fader (2005). Cíl daného modelu přesně odpovídá potřebám naší práce. Pareto/NBD řeší následující úkoly: zjišťuje pravděpodobnost, že zákazník stále žije, která je založena na transakční aktivitě v průběhu sledovaného období a odvádí očekávaný počet budoucích transakcí pro náhodně vybraného zákazníka, který je také závislý na minulém chování. Na základě znalostí o životním cyklu zákazníka (Schmitt, 2003) očekáváme, že dříve či později nastane fáze úpadku a z různých důvodů zákazník zmizí z našeho obchodního procesu: přestane nakupovat nebo odejde ke konkurenci, dokonce i může dojít k jeho likvidaci. V daném okamžiku považujeme klienta za neaktivního a tím je myšleno, že pro naši společnost už „nežije“. A proto určujeme transakční aktivitu a mluvíme o pravděpodobnosti, že zákazník ještě žije, což je velmi důležité pro analýzu chování klienta.

Konkrétní výběr také ovlivňují skutečná data, která máme k dispozici, vysoké ohodnocení a pozitivní posudky známých společností, používající daný model v praxi.

Výchozí datový soubor obsahuje informace o četnostech objednávek a časových okamžiků objednání pro skupinu všech zákazníků firmy. Naším záměrem je rozhodnout o budoucím chování dané skupiny zákazníků na základě minulých pozorování. Většina modelů, které řeší podobný problém, vychází z algoritmu seskupení zákaznických transakcí za 52 týdnů (jeden rok) a pak převádějí očekávání na individuální úroveň.

Pareto/NBD je prvním modelem, který byl zpracováván a aplikován Schmittleinem v roce 1987. Tento statistický model popisuje opakující se objednávky a okamžiky odchodu klienta z obchodního cyklu za podmínek stochastického pravidelného nákupu. Tímto je myšleno, že v průběhu určitého období zákazník uskuteční řadu transakcí a pak zmizí. Okamžik zmizení (zákazníka považujeme za neaktivního) a jeho predikce jsou pro nás hlavními objekty zájmu. Odchod zákazníka je popsán Paretovým pravděpodobnostním rozdělením, které představuje spojení exponenciálního a gama rozdělení. Chování a opakující se aktivita zákazníka je popsána pomocí negativního binomického rozdělení (zkratka z anglického jazyka NBD – negative binomical distribution). NBD představuje syntézu Poissonova a gama distribučního pravděpodobnostního rozdělení. Odtud pochází i název modelu. Pareto/NBD představuje sofistikovaný postup přípravy dat a občas vyžaduje náročné výpočty na etapě odhadu parametrů modelu. Ale v našem případě máme všechny potřebné údaje k aplikaci na reálných datech, takže se můžeme zaměřit na výše uvedený model.

Podle článku od Americké univerzity managementu a operačního výzkumu (Jerath, Fader, Hardie, 2011), model Pareto/NBD je založen na pěti předpokladech:

1. v průběhu aktivní doby klienta a při počtu transakcí x za časový úsek t , má Poissonovo pravděpodobnostní rozdělení se střední hodnotou λt :

$$P(X(t) = x | \lambda) = \frac{(\lambda t)^x e^{-\lambda t}}{x!}. \quad (2.4)$$

Rozdělení délky intervalu mezi náhodně se vyskytujícími transakcemi vyjadřuje exponenciální rozdělení, kdy pravděpodobnost výskytu jednotlivé transakce má Poissonovo rozdělení:

$$f(t_j - t_{j-1} | \lambda) = \lambda e^{-\lambda(t_j - t_{j-1})}, t_j > t_{j-1} > 0, \quad (2.5)$$

kde t_j je okamžik j -té transakce.

2. heterogenita míry uskutečněných transakcí λ popsána pomocí gama rozdělením s parametry r, α :

$$g(\lambda|r, \alpha) = \frac{\alpha^r \lambda^{r-1} e^{-\lambda\alpha}}{\Gamma(r)}. \quad (2.6)$$

3. Každý zákazník má určitou délku celoživotní hodnoty τ . Konkrétní hodnotu, když klienta považujeme za neaktivního, můžeme popsat exponenciálním pravděpodobnostním rozdělením se střední hodnotou μ :

$$f(\tau|\mu) = \mu e^{-\mu\tau}. \quad (2.7)$$

4. heterogenitu, či jinými slovy různorodost míry odchodu zákazníka popíšeme gama rozdělením s parametry s, β :

$$g(\mu|s, \beta) = \frac{\beta^s \mu^{s-1} e^{-\mu\beta}}{\Gamma(s)}. \quad (2.8)$$

5. Jak míra transakcí λ , tak i míra odchodu μ jsou odlišné a nezávislé v rámci chování od zákazníka k zákazníkovi.

Předpoklady ve vzorcích (2.4), (2.5) nám umožňují sestavit NBD model, který popisuje rozdělení počtu transakcí, v průběhu „živé“ doby chování zákazníka

$$P(X(t) = x|r, \alpha) = \int_0^\infty P(X(t) = x|\lambda) g(\lambda|r, \alpha) d\lambda = \frac{\Gamma(r+x)}{\Gamma(r)x!} \left(\frac{\alpha}{\alpha+t}\right)^r \left(\frac{t}{\alpha+t}\right)^x. \quad (2.9)$$

Zatímco předpoklady (2.6), (2.8) nám poskytují model Paretova rozdělení:

$$f(\tau|s, \beta) = \int_0^\infty f(\tau|\mu) g(\mu|s, \beta) d\mu = \frac{s}{\beta} \left(\frac{\beta}{\beta+\tau}\right)^{s+1}, \quad (2.10)$$

a

$$F(\tau|s, \beta) = \int_0^\infty F(\tau|\mu) g(\mu|s, \beta) d\mu = 1 - \left(\frac{\beta}{\beta+\tau}\right)^s. \quad (2.11)$$

V tomto okamžiku jasně vidíme, odkud se objevil ten název daného modelu Pareto/NBD. Tento model vyžaduje několik důležitých údajů o zákazníkovi: okamžik poslední objednávky/transakci a četnost/počet transakcí za sledované období. Takže už teď vidíme jednu zásadní výhodu modelu Pareto/NBD – příprava reálných dat by neměla být příliš časově a objemově náročná. Vektor, popisující tuhle informaci, vypadá následovně:

$$(X = x, t_x, T),$$

kde x je počet transakcí za sledované období $(0, T]$;

t_x – okamžik poslední realizované transakce $(0 < t_x \leq T)$;

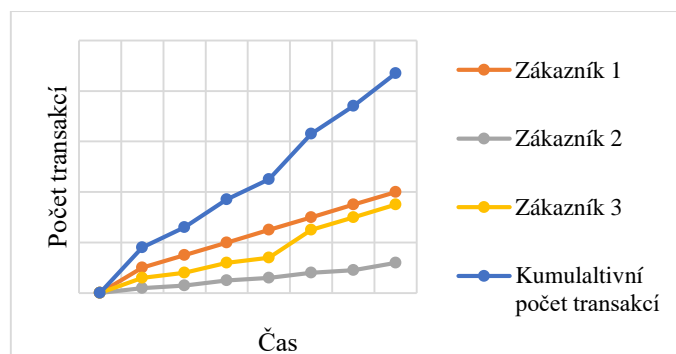
T – celková délka sledovaného období.

Model Pareto/NBD, využívající tyto klíčové statistiky, má následující parametry:

- $E[X(t)]$ – očekávaný počet transakcí jednotlivých zákazníků za sledované období délky t . Tato hodnota je velmi důležitá z hlediska výpočtu celkového očekávaného objemu prodeje v průběhu času;
- $P(X(t) = x)$ – pravděpodobnost, že náhodná veličina X , závislá na t , nabude hodnoty x , jinými slovy pravděpodobnost výskytu x transakcí za sledované období t ;
- $E(Y(t)/X = x, t_x, T)$ – očekávaný počet transakcí za období $(T, T + t]$ pro jednotlivého zákazníka s vyskytnutým chováním, popsaným vektorem $(X = x, t_x, T)$.

Věrohodnostní funkce (likelihood), zapsaná pro model Pareto/NBD, je složitá a zahrnuje výpočetní složky z Gaussova hypergeometrického pravděpodobnostního rozdělení. Kromě toho, pro interpretaci a hlubší porozumění podstaty výstupů Gaussova hypergeometrického rozdělení musíme analyzovat výsledky s ohledem na marketingovou složku CRM analýzy. Další věc, kterou musíme zmínit, je silná závislost věrohodnostní funkce na konkrétních výchozích parametrech a výpočetní náročnost. Finálním krokem maximálně věrohodného odhadu je najít bod nulové derivace, tj. bod, ve kterém věrohodnostní funkce nabývá maximální hodnoty. Problémy mohou nastat v okamžiku, kdy se budeme snažit optimalizovat numerické výpočty a přizpůsobit věrohodnostní funkci reálným datům. Takže danou skutečnost považujeme za nevýhodu.

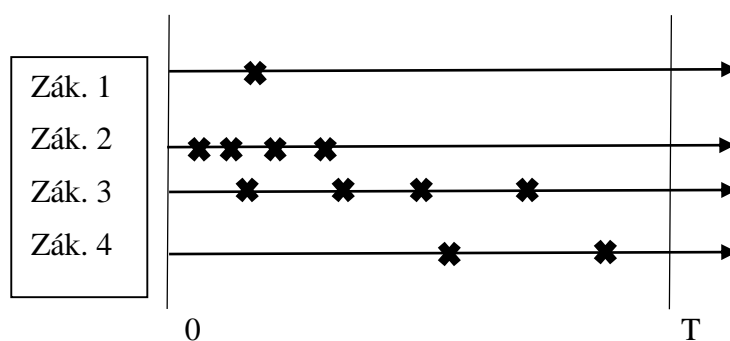
Jak uvádí ve své literatuře Fader a kol. (2005), v roce 2004 Reinartz a Kumar publikovali report, ve kterém popsali pozitivní zkušenost s pokusem využití standardní techniky maximálního věrohodného odhadu pro stanovení parametrů modelu. Alternativní způsob třístupňového odhadu parametrů byl navržen Schmittleinem a Petersonem, který byl posouzen jako jednodušší než maximálně věrohodný odhad, ale stále výpočetně náročný. Odhad parametrů modelu Pareto/NBD vyžaduje speciální software, což v našem případě je programovací jazyk R, pomocí kterého můžeme realizovat náročný matematický výpočet přes několik příkazů. Balíček v programu R, který umožňuje provést dané modelování, se nazývá BTYD (zkratka z angličtiny – Buy Till You Die). O něm budeme mluvit v další části práce. Nyní zmíníme, že tento balík je založen na principu agregace kumulativního nákupu, ale později algoritmus vede k individuálnímu sledování.



Obrázek 2.3 Agregace kumulativního počtu transakcí

Zdroj: Fader, Hardie – Journal of Interactive Marketing 23 (2009) 64

Každý zákazník má svůj vlastní bod, ve kterém odejde nebo jinými slovy zemře. Tento předpoklad je zachycen na obrázku 2.3. Ztráta zákazníka může být způsobena různými faktory, jako jsou finanční okolnosti, změna geografické polohy nebo špatné minulé zkušenosti. Ale ta příčina nás až tak moc nezajímá, důraz je kladen především na správné namodelování okamžiku odchodu a délku aktivního chování.

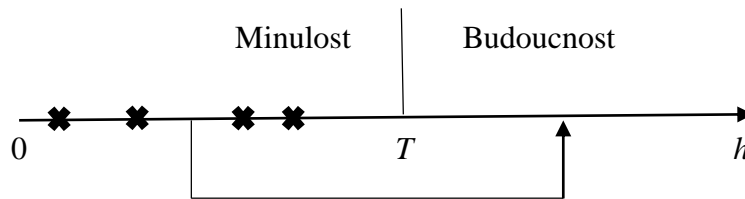


Obrázek 2.4 Příklad transakčního chování jednotlivých zákazníků

Zdroj: Fader, Hardie – Journal of Interactive Marketing 23 (2009) 64

Na dalším obrázku jsou zobrazeny příklady zákaznického chování v průběhu časového úseku $(0, T]$. Vidíme čtyři různé transakční historie. První zákazník objednával pouze jednou, zatímco druhý klient objednával čtyřikrát, ale jenom na začátku sledovaného období. Třetí případ znázorňuje pravidelné nakupování a čtvrtý klient zase nakupoval dvakrát za období. Jak je patrné z obrázku 2.4, každý klient se chová velmi odlišným způsobem. Ačkoliv zákazníci 2 a 3 mají stejný počet transakcí v daném intervalu $(0, T]$, od třetího zákazníka můžeme očekávat větší aktivitu v budoucnu. Zákazník číslo 2 už dlouhou dobu nekupoval, a pravděpodobně ho považujeme za neaktivního. Primární objekt zkoumání jsou zákazníci, kteří jsou pravděpodobněji „naživu“ v okamžiku T než ostatní a dáváme tedy přednost těmto klientům. Realita je taková, že nevíme, v jaký okamžik očekávat nákup, stejně jako nevíme moment

odchodu klienta (okamžik, kdy se stane neaktivním). Počet objednávek nebo počet transakcí jsou náhodné veličiny. Zákaznické chování analyzujeme a na základě minulých pozorování sestrojíme model, pomocí kterého odhadneme budoucí vývoj transakcí. Aplikace modelu na reálná data umožňuje získat sadu kvantitativních výstupů: pravděpodobnost, že klient v daném okamžiku ještě žije (tj. je aktivní a nakupuje), očekávaný počet transakcí v budoucím časovém horizontu a nakonec celoživotní hodnotu zákazníka. Všechny výstupy jsou založeny na předcházejícím chování klientů.



Obrázek 2.5 Proces modelování a predikce transakčního chování

Zdroj: Fader, Hardie – Journal of Interactive Marketing 23 (2009) 62

Na obrázku 2.5 sledujeme nulový bod, který je začátkem období, T – bod současnosti, oddělující minulost a budoucnost a h – horizont predikce. Takže vidíme, že závěry o budoucím chování jednotlivého klienta odvozujeme z chování současného, které je popsáno exponenciálním a Paretovým pravděpodobnostním rozdělením. Je to hlavní důvod, proč jsme se podrobně věnovali rozdělením počtů transakcí a délky intervalů mezi jednotlivými transakcemi. Možnost predikovat chování zákazníka na individuální úrovni je základem pro výpočet celoživotní hodnoty klienta.

Předpokládáme, že víme datum jednotlivých transakcí v průběhu období $(0, T]$. Na obrázku 2.5 jsme označili tento okamžik křížkem. Existují dva možné způsoby, jak tento scénář vývoje může pokračovat dál:

1. zákazník je naživu na konci sledovaného období (tj. $\tau > T$), v tomto případě individuální věrohodnostní funkce je pravděpodobnostní funkcí exponenciálního rozdělení;

$$L(\lambda|t_1, \dots, t_x, T, \tau > T) = \lambda e^{-\lambda t_1} \lambda e^{-\lambda(t_2-t_1)} \dots \lambda e^{-\lambda(t_x-t_{x-1})} e^{-\lambda(T-t_x)} = \lambda^x e^{-\lambda T}, \quad (2.12)$$

kde $t_1 \dots t_x$ jsou data transakcí. Ostatní proměnné byly vysvětleny v předchozí části.

2. Zákazník se stane neaktivním v nějaký okamžik τ z intervalu $(t_x, T]$. V tomto případě věrohodnostní funkce (likelihood) bude mít tvar:

$$L(\lambda|t_1, \dots, t_x, T, \tau \in (t_x, T)) = \lambda e^{-\lambda t_1} \lambda e^{-\lambda(t_2-t_1)} \dots \lambda e^{-\lambda(t_x-t_{x-1})} e^{-\lambda(\tau-t_x)} = \lambda^x e^{-\lambda \tau}. \quad (2.13)$$

Výhodou je, že v obou případech model nevyžaduje konkrétní datum jednotlivých objednávek. Můžeme klidně nahradit t_1, t_2, t_x, T za (x, t_x, T) , kde parametr t_x je roven nule v okamžiku, kdy x je rovno nule. Jinými slovy t_x a x jsou minimálními postačujícími statistikami zákaznické transakční míry. Jeden ze způsobů, jak můžeme posoudit kvalitu odhadu parametrů, je zjistit rozptyl odhadu. Nejmenší rozptyl umožňuje nahradit celý výběr jednou jedinou statistikou, která obsahuje veškerou informaci o jednotlivých parametrech. Jinými slovy, číslo představující minimální postačující statistiku je natolik reprezentativní, že povoluje nahradit celou datovou řadu.

Čtyři odhadované parametry (r, α, s, β) stanovíme na základě metody maximální věrohodnosti. Velmi podrobné se touto problematikou a vysvětlením věrohodnosti zabývala I. Malá (Statistické úsudky, 2013). „Nejčastěji používanou metodou konstrukce odhadu je metoda maximální věrohodnosti. Nejprve je třeba definovat věrohodnostní funkci. Do sdružené hustoty náhodného výběru dosadíme hodnoty pozorování x získané náhodným výběrem a budeme ji považovat za funkci parametru. V takovém případě budeme sdruženou hustotu značit L a budeme ji nazývat věrohodnostní funkcí“. Předpokládejme, že náš datový soubor zahrnuje N zákazníků, kde i – tý zákazník má x_i transakcí za období $(0, T_i]$ a poslední transakce nastala právě v čase t_{xi} . Jak uvádí I. Malá (2013) – „Při hledání maximálně věrohodného odhadu obvykle místo maximalizace věrohodnostní funkce hledáme maximum logaritmu. Vzhledem k tomu, že věrohodnostní funkce je nezáporná, lze ji logaritmovat“. Logaritmická věrohodnostní funkce bude mít tvar:

$$\log(L) = \sum_{i=1}^N \ln[L(r, \alpha, s, \beta | x_i, t_{xi}, T_i)]. \quad (2.14)$$

Dále následuje derivace logaritmické věrohodnostní funkce podle jednotlivých parametrů:

$$\frac{\partial \ln L(\theta)}{\partial \theta} = \frac{\partial l(\theta)}{\partial \theta}, \quad (2.15)$$

kde θ je obecné značení pro odhadované parametry. „Pokud existuje derivace věrohodnostní funkce podle parametru θ , lze maximum funkce hledat jako bod, ve kterém je nulová první derivace“.

Nakonec položíme derivaci rovnou nule. Finálním krokem je řešení této rovnice a určení bodu maxima:

$$\frac{\partial l(\theta)}{\partial \theta} = 0. \quad (2.16)$$

Ačkoliv se budeme zabývat aplikací základního modelu Pareto/NBD, zmíníme existenci jiné rozšířené verze. V roce 2005 dříve zmínění autoři (Fader, Hardie, Lee) doplnili hlavní model o řadu dalších možností. Pomocí submodelu můžeme počítat jak s celkovým tokem transakcí, tak i s individuální útratou, ohodnotit je a vyčíslit finanční hodnotu, tj. průměrný

výdaj na jednu transakci. V tomto případě výzkumníci předpokládají nezávislost průměrného výdaje zákazníka na celkovém transakčním toku. Později tento předpoklad, který vyvolával různá omezení v použití, byl doplněn pohodlnějšími přístupy (Glady, Baesens, Croux, 2008).

Další variantou úpravy modelu Pareto/NBD je spojení myšlenky o průměrném výdaji na transakci s pojmem „reziduální transakce“, což je snížení očekávané objednávky o nějakou nesystematickou složku (Fader, Hardie, Lee, 2005). Tento přístup umožnil společnosti získat údaje očekávané celoživotní hodnoty zákazníka, která už také bude zahrnovat odhad nesystematické složky, založenou na minulém a skutečném chování klienta. Klíčový přínos daného postupu spočívá v objemu informace o nákupním chování zákazníka za sledované období. Stačí nám vědět lhůtu poslední aktivity klienta, frekvenci a sumu objednávek za sledované období (což není nic jiného, než RFM model). V některých případech nejsou k dispozici potřebné údaje, ale i pro tuto situaci existuje doplnění a řešení problému odhadu parametrů z limitovaných dat (Jerath, Fader a Hardie, 2011).

Takže naším hlavním záměrem v průběhu příští kapitoly bude aplikovat teoretický model Pareto/NBD na empirickém příkladu společnosti „Práce pro Vás“ a predikovat celoživotní hodnotu zákazníka.

3. Aplikace Pareto/NBD modelu na reálná data společnosti

V této části bakalářské práce je jedním z hlavních cílů vypočítat celoživotní hodnotu zákazníků firmy „Práce pro Vás“ pomocí Pareto/NBD modelu, který byl teoreticky popsán v podkapitole 3.4 a využít nástroje z oblasti marketingu pro splnění účelů práce – popsat a predikovat celoživotní hodnotu zákazníka. Chtěla jsem ukázat, že problematika zkoumaná marketéry se dá kvantifikovat, analyzovat a interpretovat. Často se stává, že oddělení marketingu má ve firmě špatnou reputaci kvůli složitému vyhodnocení výsledků práce a dochází ke střetu zájmů oddělení prodeje a marketingu. Marketéři se nejenom snaží uspokojovat potřeby zákazníka a vymyslet vhodnou strategii pro značky na trhu, ale také používají kvantitativní metody pro řízení vztahů se zákazníky. V druhé subkapitole se věnujeme propojování a integraci kvalitativních a kvantitativních metod, čímž se snažíme sdružit statistiku a marketing. Zároveň vyzkoušíme na praktickém příkladu popsany model Pareto/NBD v marketingové praxi.

Pro návrhy řešení a metody řízení vztahů se zákazníky uvedené v první části práce je třeba definovat výchozí předpoklady. Jakákoliv analýza začíná z vnitřního auditu podniku, který může zahrnovat různé složky v závislosti na problému, který řešíme.

3.1 Krátký popis analyzované společnosti

S.r.o. „Práce pro Vás“ je běloruská společnost, která je jedním z lídrů na běloruském trhu novin o zaměstnání a vzdělání. Podnik provozuje svou činnost od roku 1998, ve kterém byl založen. Periodicita tisku je třikrát týdně (pondělí, středa a čtvrtek). Od roku 2008 firma poskytuje služby nejenom v hlavním městě Minsku, ale i na republikánské úrovni ve všech oblastních centrech Běloruska. Celková týdenní tiráž novin je 42.000 kopií. Tisk má také své webové stránky <http://rdw.by/>, kam klienti inzerují (doplňk k novinové inzerci), což umožňuje zachytit celý segment potenciálních uchazečů o zaměstnání, od dělníka továrny po vedoucího oddělení.

Společnost „Práce pro Vás“ provozuje činnost jak pro právnické osoby na trhu B2B (jde o inzerování informací o nabídkách práce), tak i pro konečného spotřebitele na trhu B2C, který kupuje noviny s účelem hledání volného pracovního místa. Ale v další částech práce budeme zkoumat a analyzovat především prostředí B2B trhu. Činnost se týká především prodeje inzertních ploch v samotných novinách a na webových stránkách firmy. Firma realizuje prodej na fakturu (bez předběžné úhrady platby), a proto pracovníci musejí řešit pravidelně vznikající problém krátkodobých a dlouhodobých pohledávek.

K základním procesům obchodního oddělení patří:

1. získávání nových zákazníků;
2. práce se stálými zákazníky;
3. monitorování rizika, práce s dlužníky;
4. vymáhání pohledávek, mladších než 1 měsíc;
5. vystavení splátkových kalendářů.

Často dochází ke střetu zájmů finančních služeb společnosti s obchodním oddělením, které se snaží více inzerci prodat, ale nemůže znovu začít poskytovat služby, než budou uhrazeny staré pohledávky. Mezi hlavní faktory, které ovlivňují odvětví firmy, jsou sezónnost, nízká míra nezaměstnanosti, odliv pracovních sil, rychle měnící legislativa na trhu masmédií, vysoké náklady na tisk, nestabilní inzerování a odliv klientů kvůli ekonomické krizi. Tato realita vyžaduje kvalitní analýzu chování klientů, modelování a predikci tohoto chování, minimalizaci sezonní složky a snahu udržet dlouhodobý vztah s klíčovými klienty firmy.

3.2 Popis řešeného problému

Pro spoustu podniků se hlavní podnikatelský záměr pohybuje kolem snahy poznat klienta, kterého musí držet za každou cenu, a kterého ne. Dnešní společnosti se setkávají s rostoucí konkurencí a globalizací. Dnes je důležité být lepší a rychlejší než konkurenti, je nutné odlišit cílového zákazníka, rychleji reagovat na potřeby klienta, na podněty trhu a nacházet každý den nové řešení. Je třeba pečovat o významné zákazníky spíše více než o zákazníky „okrajové“, jejichž nákupy jsou náhodné, bez určení dlouhodobých vzájemných kontaktů. Kvůli prudkému vývoji informačního prostředí, rozvoje konkurence na trzích, rychlých změn v podnikatelském prostředí nemá žádná organizace záruku neměnné struktury zákazníků. Z tohoto faktu je zřejmé, jak je důležité v současných podmínkách se věnovat zákazníkům, diferencovat a segmentovat přístupy k nim, pravidelně vyhodnocovat, kontrolovat a korigovat firemní chování a péči o zákazníka. Tradiční přístupy k marketingu a marketingové komunikace se jeví jako nedostatečné. Manažeři vyžadují stále nové metody, modely a druhy šetření v této oblasti. Tato skutečnost způsobila nastavení celého podniku tak, aby byly osloveni a udrženi důležití zákazníci. Zákazníci jsou jádrem a bodem, ve kterém se propojují všechny zdroje, které jsou podniku dostupné.

Takže z výše popsané reality můžeme stanovit konkrétní problémy:

- analýza a predikce celoživotní hodnoty zákazníka na reálných datech;
- výběr vhodného modelu pro simulaci celoživotní hodnoty;

- příprava dat o zákaznicích pro potřeby modelu Pareto/NBD;
- aplikace modelu Pareto/NBD na upravená data;
- určení základních charakteristik klíčových klientů, které inzerují pravidelně a predikce jejich chování.

V průběhu následující podkapitoly se zaměříme na řešení stanovených problémů, pak především na modelování zákaznického chování a dále na udržování dlouhodobých vztahů se zákazníky, což je cílem naší práce.

3.3 Popis a příprava reálných dat

Začneme analýzu jedním z nejvýznamnějších kroků – popíšeme a připravíme data pro potřeby modelu Pareto/NBD (McCarthy, Wadsworth a kol., 2014). První etapa přípravy dat je úprava skutečné klientské databáze pomocí tabulkového procesoru. Primární data obsahují 11 sloupců, mezi kterými jsou unikátní identifikační číslo klienta (dále bude používaná zkratka ID), identifikační číslo výtisku, datum objednávky, typ inzerátu (modul nebo řádková inzerce), velikost modulu, zvolená rubrika a sleva v procentech. Bohužel nejsou k dispozici ceny objednávek, budeme je tedy dopočítat „ručně“ v Excelu pomocí funkcí „index“ a „pozvyhledat“. Uvedené funkce umožňují vyhledat konkrétní výraz z jiné tabulky. S použitím těchto funkcí doplníme tabulku sloupcem cen beze slevy a pak dopočítáme výslednou cenu s poskytnutou slevou. Dále jsme odstranili chybějící pozorování a objednávky, které měly slevu víc než 95 procent (sleva byla poskytnuta sociálním a neziskovým organizacím). Extrémní hodnoty mají negativní vliv na základní charakteristiky klientského chování a zkreslují realitu. Posledním krokem je přepočítání ceny objednávky na české koruny.

Tabulka 3.1 Fragment průběžné přípravy datového souboru

ID Klient	ID výtisk	Datum	Typ inzerátu	Velikost	Rubrika	Sleva	Cena beze slevy, BYR	Cena včetně slevy, BYR	Cena beze slevy, CZK	Cena včetně slevy, CZK
754	1990	01.04.2016	Modul	M1/3	0-první	-63,2	935 927	344 421	1 123	413
1134	1990	01.04.2016	Modul	M1/2	00-poslední	-50	455 803	227 902	547	273
2052	1990	01.04.2016	Modul	M1/2	7-maloobchod	-65	260 030	91 011	312	109

Zdroj: vlastní výpočet

Celkem data v Excelu obsahují 13 461 řádků (počet objednávek) za období od 5. 1. 2015 až do 1. 4. 2016, v tabulce 3.1 je ukázka z tohoto souboru dat. Zvolený interval se skládá ze 448 dnů,

což je 64 týdnů. Pro další zpracování a potřeby modelu Pareto/NBD upravíme frekvenci dat na týdenní. V průběhu této doby služby byly poskytnuty 2 138 unikátním klientům.

Tabulka 3.2 Počet opakujících se objednávek

Počet nákupů	Počet unikátních zákazníků
1	1053
2	316
3	214
4	97
5	82
6	76
7	53
8	35
9	32
10	28
11	20
12	18
13	17
14	15
15	11
16	7
17	7
18	6
19	6
20	5
21	5
22	8
23	5
24	2
25	4
26	2
27	2
28	1
30	2
31	2
33	1
35	2
39	1
44	1
45	1
49	1
Celkem	2138

Tabulka 3.2 zobrazuje přehled, kolik společnost „Práce Pro Vás“ má vracejících se zákazníků. První sloupec obsahuje počet nákupů nebo jinými slovy počet transakcí, a druhý sloupec zobrazuje počet unikátních zákazníků s daným počtem transakcí. Vzhledem k tomu, že v další části budeme pracovat s týdenními daty, tak pod počtem opakujících se transakcí rozumíme počet týdnů, ve kterých konkrétní zákazník provedl alespoň jednu transakci. Z tabulky 3.2 vyplývá, že maximální počet opakujících se objednávek je 49. Tím je myšleno, že jeden zákazník opakovaně kupoval služby firmy po dobu 49 týdnů. A naopak z prvního řádku tabulky vidíme, že 1 053 unikátních zákazníků nakupovalo pouze jednou. Suma druhého sloupce je rovna 2 138 (což je celkový počet unikátních zákazníků firmy za sledované období). Pomocí dané tabulky je dobře vidět, že počet opakování se prudce snižuje s každým novým nákupem a na konci tabulky vidíme pouze jednotlivce. Na základě dat z této tabulky jsme získali velmi zajímavý a užitečný výstup: můžeme říct, že z 2 138 klientů se k nám vrátilo pro opakovanou objednávku 1 085 zákazníků (od celkového součtu druhého sloupce jsme odečetli počet klientů s jednou objednávkou), což je více než 50 % původního počtu klientů. Ale rozdíl mezi jedním nákupem a dvěma nákupy je příliš výrazný, z dlouhodobého hlediska je to negativní efekt, což ještě jednou potvrzuje nutnost analyzovat a predikovat CLV. Pro základní kvantitativní proměnné uvedeme hlavní ukazatele popisné statistiky. Pro lepší představu jsou údaje představeny ve formě následující tabulky. Střední hodnota slevy je 23,35 %, přestože modus dané proměnné je nulový. Víme, že průměr je velmi citlivý na extrémní hodnoty, a proto

větší pozornost věnujeme mediánu a modální hodnotě. Medián má hodnotu 20 procent, tato

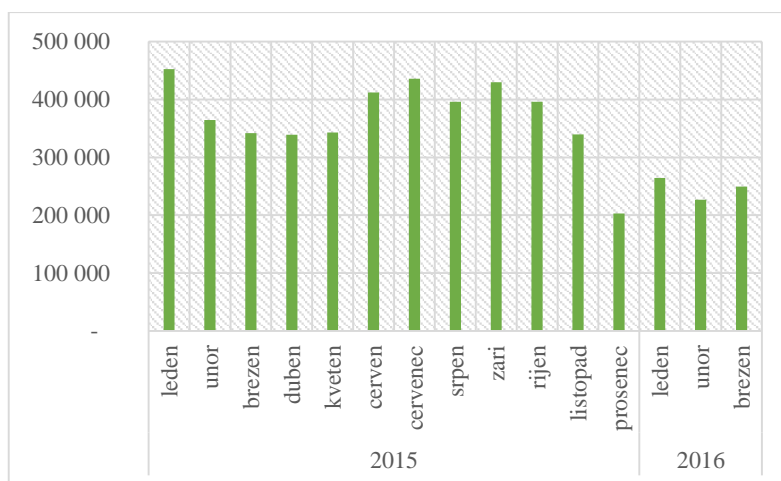
hodnota rozděluje náš datový soubor na dvě stejně velké části. Celkové tržby za období jsou 7.555.854 korun, ale včetně slevy je tento ukazatel o 2.355.447 korun menší, což lze považovat za ušlý zisk. V budoucnu pomocí provedené analýzy a výstupů z modelu Pareto/NBD můžeme navrhnout dané firmě věnovat větší pozornost cenové politice a poskytovat slevy výhradně cílovým klientům s delší očekávanou celoživotní hodnotou pro udržování dlouhodobých vztahů. Uvedené míry variability, jako jsou rozptyl, jeho odmocnina, třetí a čtvrtý centrální momenty, samy o sobě užitečnou informaci nenesou, ale jsou vhodné pro porovnání navzájem. Pomocí hodnot minima, maxima a variabilního rozpětí jsme schopni poznat, jak různí jsou naši zákazníci jsou a jaké různorodé služby z hlediska ceny jsou jim poskytovány. Musíme dodržovat Paretovo pravidlo: 80% zisku přináší 20% klientů. Tento fakt potvrzuje nutnost pečlivé průběžné analýzy, cílového přístupu k zákazníkům a udržování dlouhodobých vztahů.

Tabulka 3.3 Popisná statistika základních kvantitativních proměnných

	Sleva (v %)	Cena bez slevy, CZK	Cena včetně slevy, CZK
Stř. hodnota	-23,3	565,8	389,4
Chyba stř. hodnoty	0,2	5,8	3,9
Medián	-20	312	252
Modus	0	312	126
Směr. odchylka	25,5	674	446,8
Rozptyl výběru	649,3	454 230	199 658,5
Špičatost	-0,5	16,8	20
Šikmost	-0,8	3,4	3,6
Rozdíl max-min	95	6 619,4	4 917,9
Minimum	-95	48	6,5
Maximum	0	6 667,4	4 924,4
Součet		7 555 854	5 200 407

Zdroj: vlastní výpočet

Dále uvedeme graf průměrných měsíčních tržeb za celé sledované období od ledna 2015 do března 2016. Graf 3.1 je užitečný jak k porovnání měsíčních sum transakcí, tak i prvních čtvrtletí roků 2015 a 2016. Například je dobře vidět, že leden roku 2015 byl mnohem ziskovější než stejný měsíc roku 2016. Tento pokles tržeb ze 452 tisíc korun na 264 tisíc způsobila ekonomická krize v Bělorusku a sousedních státech. Stejný trend vidíme i v únoru a březnu – tržby klesají, ale chování zákazníka se vyvíjí shodným způsobem: v lednu se prodává největší objem služeb, pak počet objednávek pomalu klesá. Z grafu je také viditelná sezonní složka, která se vyskytuje pravidelně na pracovním trhu. V létě sledujeme postupný nárůst sumy transakcí, protože většina zaměstnanců má nárok na dovolenou, a proto firma musí hledat náhradní lidské zdroje.



Obrázek 3.1 Průměrné tržby za sledované období

Dále následuje druhá etapa přípravy dat, v průběhu které upravíme tabulku pro potřeby modelu Pareto/NBD a programovacího jazyka R. Nejprve nainstalujeme balíček BTYD (zkratka z anglického jazyka – Buy Til You Die, lze si ji vyložit jako sledování nákupů klientů, až do okamžiku, kdy zmizí z obchodního cyklu firmy), který zahrnuje i jiné modely, ale my se soustředíme na Pareto/NBD. Daný model je určen zejména pro nesmluvní vztahy, kdy klient může objednávat nepravidelně v jakýkoliv okamžik. Pareto/NBD popisuje a simuluje okamžik pořízení služby a moment odchodu z našeho okolí, jinými slovy analyzuje celoživotní hodnotu zákazníka. Jednou ze zjevných výhod modelu je počet parametrů. Nepotřebujeme více než čtyři proměnné. Pro každého zákazníka určíme počet transakcí za sledované období, okamžik poslední transakce a celkovou dobu pozorování. Matice, používaná v modelu, se jmenuje „customer by sufficient statistic matrix“, ale z důvodu častého použití v celé práci ji definujeme jako tzv. datovou matici, kde řádek odpovídá určitému objednateli a ve sloupcích se nachází jednotlivé proměnné. Před sestavováním modelu musíme seřadit data v R chronologicky a sloučit všechny transakce, které byly provedeny ve stejný den. Ve výsledku jsou data seřazená a každý řádek odpovídá konkrétnímu zákazníkovi s celkovou sumou objednávek za kalendářní den. Počet pozorování se zmenšil na 7 008 řádků (původně to bylo 13 461). V tuto chvíli už jsou reálná data připravena pro analýzu a sestavování modelu Pareto/NBD.

3.4 Aplikace modelu Pareto/NBD na reálná data pomocí R

V této podkapitole využijeme připravená data, odprezentujeme výsledky modelování graficky, nasimulujeme predikce celoživotní hodnoty zákazníka a popíšeme jeho chování.

Prvním krokem ověříme, jestli model funguje správně. Kontrola funguje na principu, že časovou řadu rozdělíme na dvě části: první část (trénovací) slouží k výběru modelu, odhadu jeho parametrů a ověření vhodnosti pomocí interpolačních kritérií, druhá část (validační) se používá k určení předpovědi známé skutečnosti (prognózy "ex post") a pro ověření jejich přesnosti. Datový soubor rozdělíme v poměru 70:30, necháme delší trénovací období na úkor validačního z důvodu přesnějšího modelování, navíc takto získáme spolehlivější parametry a důvěryhodný model predikce CLV. Sledována data rozdělíme v okamžiku 15. 11. 2015, který rozdělí náš datový soubor v potřebném poměru. Tedy trénovací období trvá 45 týdnů a validační období je 19 týdnů, celkem máme data za 64 týdnů.

Druhý krok je sestavení matice, která bude obsahovat počet objednávek nebo sumu celkové objednávky každého zákazníka v konkrétní den. Vycházíme z toho, která matice obsahuje pro nás více informací. V případě firmy „Práce pro Vás“ se často stává, že hodně klientů inzeruje ve stejný den, který předchází dnu tisku novin (například výtisk je připraven ve středu, a proto většina klientů objednává inzerci v úterý). Takže více „informativní“ bude pro nás bude matice s celkovou sumou objednávek zákazníka za určitý den.

Nyní můžeme sestavovat datovou matici, kterou jsme popsali výše, kde řádek odpovídá určité objednávce a sloupce jsou tvořeny třemi proměnnými – počtem transakcí za sledované období, okamžikem poslední transakce a celkovou dobou pozorování. Použijeme první sestavenou matici, obsahující sumu celkové objednávky každého klienta za den, začátek a konec sledovaného období a okamžik konce validační části. Což znamená, že potřebujeme vybrat periodicitu sledování. Pro lepší znázornění a orientaci ve výsledcích vybereme týdenní frekvenci: celkem náš datový soubor zahrnuje 64 týdnů.

Ted' už máme data v korektním formátu a následuje třetí zásadní krok – odhad parametrů modelu Pareto/NBD. Naším úkolem je odhadnout neznámé veličiny v závislosti na pozorovaných datech. Odhad sestává ze dvou základních částí: formulace modelu, který popisuje realitu a ověření shody našeho modelu se skutečným chováním zákazníka na základě reálných dat.

S ohledem na to, že jsme průběžně připravovali datový soubor k použití v programu R – odhad parametrů provedeme jedním jediným příkazem, který vyžaduje trénovací část z druhé sestavené matice, sloužící pro ověření a kalibraci modelu. Stanovili jsme čtyři parametry: r , α , s , β , které jsou uvedeny v tabulce 3.4.

První kvantitativní výstup, který jsme získali z modelu, tedy vypadá následovně:

Tabulka 3.4 Odhad parametrů modelu Pareto/NBD

r	α	s	β
0,694	6,518	0,371	9,051

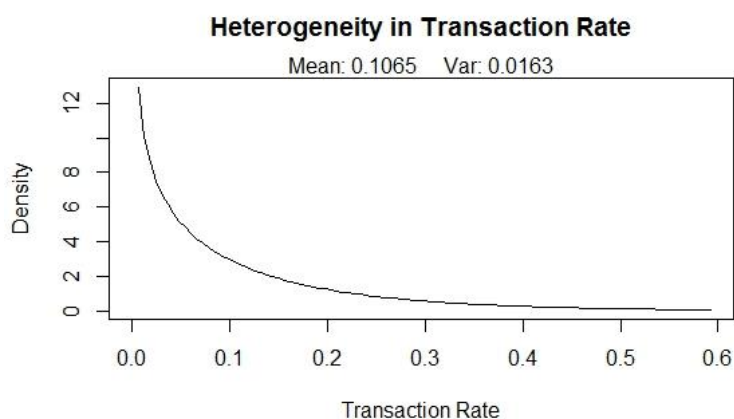
Zdroj: vlastní výpočet v R

Dále uvedeme ještě hodnotu maximalizační věrohodnostní funkce (L), která je součástí metody maximální věrohodnosti. Maximální věrohodný odhad je asymptotický nezkreslený, konzistentní a asymptoticky vydatný. Často se používá logaritmická věrohodnostní funkce, která vrátí $\log(L)$ pro soubor odhadnutých parametrů. V našem případě hodnota maximalizační funkce L je rovna -11 165,961. Ale je to jen prvotní výpočet, dále musíme zkontrolovat, zda hodnoty parametrů konvergují k odhadované hodnotě. Konzistentnost odhadu je zásadní vlastnost. Malá (2013) a Marek (2012) uvádějí následující definici: „Pro konvergenci odhadu se používá konvergence podle pravděpodobnosti. Odhad X je konsistentním odhadem parametru θ , jestliže X konverguje k odhadované hodnotě θ podle pravděpodobnosti (pro n jdoucí do nekonečna). Konzistence zaručuje, že se s rostoucím rozsahem výběru budou hodnoty odhadu pohybovat v malém okolí s velkou pravděpodobností. Pokud odhad k odhadované hodnotě podle pravděpodobností nekonverguje (není konzistentní), je jeho použití pro odhad parametrů modelu velmi omezené“. Víme od autorů článku o modelu Pareto/NBD (Fader a kol., 2005), že konvergence parametrů je ověřená a s rostoucím počtem pokusů roste i přesnost odhadů. Stanovené parametry modelu konvergují k odhadu a jsou použitelné pro další zpracování (McCarthy a Wadsworth, 2014). Z těchto ukazatelů vypočteme tři důležité charakteristiky:

1. Dělením r/α získáme hodnotu 0,106. Tohle číslo odpovídá počtu transakcí za týden, které v průměru uskuteční každý zákazník.
2. Dělením s/β získáme hodnotu 0,041. Toto číslo znamená průměrnou míru zákaznického odchodu. Můžeme říct, že v průměru za sledovaný časový úsek (v našem případě jeden týden) odejde 0,041 zákazníka nebo dále spočítáme údaj, který ukáže, za kolik týdnů odejde v průměru 1 zákazník.
3. Spočítáme-li převrácenou hodnotu k číslu 0,041, obdržíme údaj, který říká, kolik týdnů bude zákazník v průměru aktivní. V našem případě $0,041^{-1}$ je 24 - zákazník bude pravidelně nakupovat služby společnosti v průměru 24 týdnů. S ohledem na to, že data, která jsme měli k dispozici za 64 týdnů, byly rozděleny v poměru 70:30, musíme

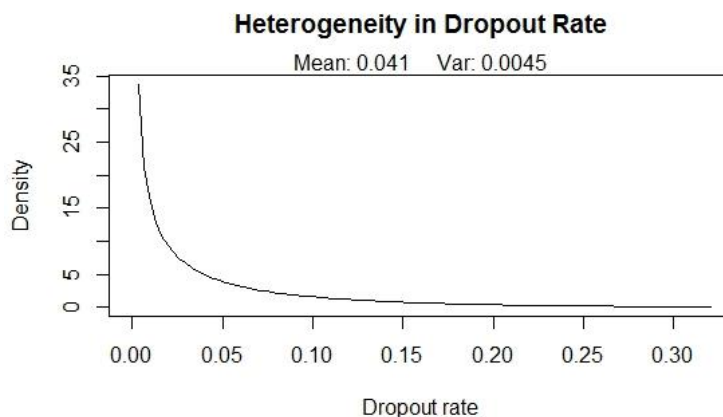
zmínit, že v trénovací části zůstalo 45 týdnů. V případě, že bychom měli delší historii transakcí, by toto číslo pravděpodobně narostlo.

Dalším krokem analýzy bude interpretace parametrů. Jak víme z minulé subkapitoly, kde jsme se podrobně zabývali vybranou statistickou metodou, parametry r a α popisují zákaznické transakce pomocí distribučního pravděpodobnostního rozdělení gama. Toto pravděpodobnostní rozdělení je vidět na následujícím grafu:



Obrázek 3.2 Stanovení parametrů gama rozdělení modelu Pareto/NBD

Graf 3.2 znázorňuje souhrnnou tendenci k nakupování pomocí předpokladu modelu Pareto/NBD, že proces individuálního nakupování se dá modelovat pomocí Poissonova pravděpodobnostního rozdělení. Na ose X vidíme míru frekvence nebo také počet transakcí za týden, který uskuteční každý zákazník, osa Y zobrazuje hustotu pravděpodobnosti samotné frekvence transakcí. Osa X odpovídá ukazateli r/α , který jsme vypočítali z tabulky 3.5. Takže můžeme sledovat, že jde tady opravdu o gama rozdělení. Z grafu lze také vyčíst, že parametry s a β patří k exponenciálnímu gama rozdělení, které popisuje Paretův proces odchodu zákazníků. V grafu je zobrazena i průměrná hodnota proměnné frekvence transakcí r/α , kterou jsme stanovili výše (0,106), rozptyl sledované veličiny je 0,0163. Rozptyl se lépe hodí k vzájemnému porovnání veličin. Následující obrázek 3.3 znázorňuje tendenci ke zmizení z nákupního procesu. Na ose X opět sledujeme počet zákazníků, kteří odejdou během sledovaného období, osa Y zobrazuje hustotu pravděpodobnosti odchodu. Střední hodnotu už známe z tabulky 3.5 a můžeme si ji „nechat“ potvrdit hodnotou z grafu 3.3 – je to 0,041, rozptyl je 0,0045. Vidíme, že variabilita míry frekvence transakcí je větší než u míry odchodu zákazníka. Ale na obou grafech je viditelný společný trend: zákaznické chování odpovídá nižším hodnotám.



Obrázek 3.3 Stanovení parametrů exponenciálního gama rozdělení modelu Pareto/NBD

V daném okamžiku máme čtyři parametry základního souboru, a můžeme přistoupit k odhadu parametrů pro klienty na individuální úrovni. Nejprve odhadneme počet očekávaných transakcí od zákazníků za sledovaný časový úsek, například za validační období, které máme dlouhé 19 týdnů, na základě údajů z trénovacího období. V průběhu minulých kroků, včetně odhadu parametrů modelu, jsme se rozhodli pro přehlednost používat týdenní periodicitu. Pomocí speciálního příkazu, který existuje v balíčku BTYD právě pro tento účel, získáme první individuální parametr. Z výstupu v R jsme stanovili, že za 19 týdnů můžeme očekávat přibližně 1,6 transakcí od našich zákazníků. Tento časový úsek odpovídá validačnímu období. Kromě toho stejným způsobem můžeme získat očekávané údaje za konkrétního zákazníka, který nás nejvíce zajímá. Následující údaje budou podmíněny minulým nákupním chováním během trénovacího období. Nyní stanovíme počet skutečných transakcí v první části našeho datového souboru a pravděpodobnost, že zákazník je ještě naživu na konci tohoto trénovacího období. A opět využijeme týdenní periodicitu, podle které jsme odhadovali parametry modelu. Vybereme libovolného klienta, například klienta s identifikačním číslem „1481“. Z vytvořené datové matice můžeme zjistit individuální charakteristiky daného zákazníka, které jsou zobrazeny v následující tabulce.

Tabulka 3.5 Charakteristiky konkrétního klienta s ID 1481

X	$t. x$	$T. cal$
12	34,286	43,857

Zdroj: vlastní výpočet v R

Kde:

x – počet objednávek za trénovací období,

t, x – pořadí týdnu, ve kterém se odehrávala poslední opakovaná objednávka,

T, cal – délka sledovaného trénovacího období.

Pravděpodobnost, že uvedený zákazník bude naživu na konci trénovacího období, tj. za 45 týdnů, je 0,718 a v průběhu validačního období můžeme od něj očekávat 3,24 transakcí na základě jeho minulého chování v trénovacím období. Tento průměrný očekávaný počet opakujících se transakcí můžeme stanovit i pro náhodného klienta během stejného časového období (45 týdnů). Pro tento účel autoři článku o predikci CLV a porovnání výsledků různých typů modelování (Jašek, Vraná, 2014) využili následující vzorec, který byl původně vytvořen Faderem, Hardiem a Leeem (2005):

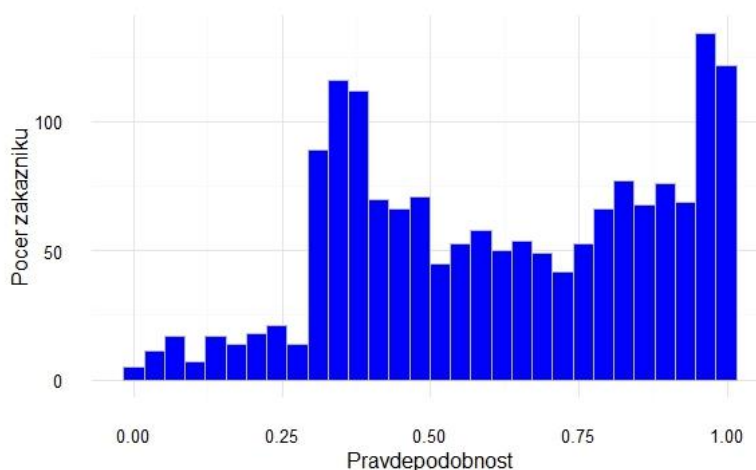
$$E[X(t) | r, \alpha, s, \beta] = \frac{r\beta}{\alpha(s-1)} * \left[1 - \left(\frac{\beta}{\beta+t} \right)^{s-1} \right] \quad (3.1),$$

kde r, α, s, β jsou stanovené výše parametry;

t – délka časového horizontu.

V našem případě toto číslo je rovno 3,18, tudíž od náhodného zákazníka v průběhu 45 týdnů můžeme očekávat v průměru 3,18 opakovaných objednávek.

Vzhledem k tomu, že už máme hodnotu pravděpodobnosti o zákazníkovi „1481“, tak můžeme přidat graf 3.4, který vizualizuje rozdělení pravděpodobnosti, že zákazníci jsou aktivní, pro celou databázi klientů. Na ose X vidíme hodnoty pravděpodobnosti a osa Y zobrazuje počty zákazníků odpovídající určité pravděpodobnosti, že zákazník je aktivní.



Obrázek 3.4 Rozdělení pravděpodobnosti aktivity zákazníků

Graf 3.4 zobrazuje celkový přehled, kolik klientů nám zůstalo naživu na konci trénovacího období s určitou pravděpodobností. Z daného grafu je zřejmé, že společnost je v docela

„dobrých číslech“. Ačkoliv hodně klientů se pohybuje kolem pravděpodobnosti 0,3 – 0,4, tak pozorujeme značný nárůst počtu klientů mezi pravděpodobnostmi 0,7 – 1.

Na dané etapě můžeme sledovat paradox rostoucí četnosti.

Tabulka 3.6 Paradox rostoucí četnosti

X	Očekávaný počet transakcí
10	5,479
15	2,601
20	0,760
25	0,180

Zdroj: vlastní výpočet v R

I když model Pareto/NBD má matematickou a empirickou podstatu a výsledky jsou věrohodné, tak občas můžeme pozorovat netypické důsledky. Jedním z nich je paradox rostoucí frekvence, který je zobrazen v tabulce 3.6, kde první sloupec je určen pro počet transakcí v trénovacím období, druhý sloupec obsahuje průměrné údaje očekávaného počtu transakcí za 45 týdnů. Podrobně se tímto problémem zabývali autoři článku (Fader, Hardie, Lee: Using Iso-value Curves for Customer Base Analysis, 2004), který popisuje příčinnou souvislost paradoxu a CLV: „Frekvence objednávek a okamžik poslední transakce značně ovlivňují celoživotní hodnotu zákazníka; v případě, že oba ukazatele mají vysoké hodnoty, pozorujeme pozitivní vliv na CLV. Čím delší je okamžik posledního nákupu, tím se CLV a potenciál nakupování snižují, i když před tím byla u klienta vysoká četnost objednávek.“

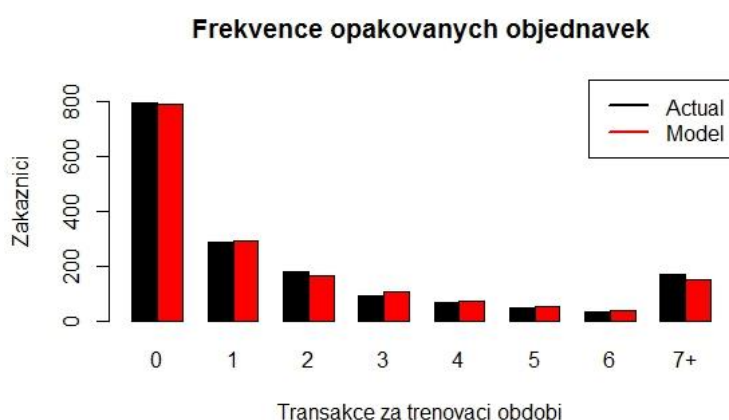
Podle stanovených podmínek modelu, větší četnost transakcí je spojená s vyšší délkou celoživotní hodnoty zákazníka. Ale u zákazníků, kteří nenakupují delší dobu, vyšší frekvence indikuje větší hodnotu věrohodnostní funkce, že klient již není aktivní (tudíž má nižší celoživotní hodnotu).

Kromě úsudku o jednotlivých zákaznících, náš model umožňuje porovnat očekávané a skutečné klientské chování pro oba časové úseky: trénovací a validační. Na grafu 3.5 vidíme, že model sedí celkem dobře a popisuje skutečná data kvalitně. Černou barvou jsou označeny skutečné údaje a červenou očekávané objednávky zákazníků. Pro lepší přehlednost přidáme ke grafu tabulku dat 3.7, ze kterých graf byl vytvořen.

Funkce pro sestavení grafu, která umožňuje generovat očekávaná data, vyžaduje několik údajů: odhadované parametry a údaje z datové matice v trénovacím období. Kromě toho potřebujeme stanovit konkrétní číslo, které se nazývá cenzor. Toto číslo nemůže být vyšší než největší četnost v trénovacím období. Výchozí struktura bude obsahovat cenzor + 1 skupin, do

kterých data roztřídíme. Toto číslo podle definice od McCarthy, Wadsworth (2014) určuje, kde data budou „uříznuta“. Jak uvádí Mann (1974) – „Ve statistickém výzkumu se cenzor používá v případě, kdy rozsah hodnot měření nebo hodnot pozorování je částečně známý. Například analyzujeme dopad drog na míru úmrtnosti. Podle datového souboru nevíme, jaký je maximální věk, ve kterém osoba zemře, buď ve věku 75, nebo starší, tj. z datového souboru nemůžeme určit, jestli konkrétní osoba je naživu v daném věku nebo ne. Z toho důvodu stanovíme číslo cenzor na úrovni 75, a tím budeme myslet, že údaje větší než 75 stejně budou zařazeny do této skupiny“.

Stanovíme cenzor na úrovni 7. Tolik tříd bude stačit pro znázornění četnosti a zároveň tím splníme podmínky použití dané funkce.



Obrázek 3.5 Frekvence opakovaných objednávek v trénovacím období

Graf 3.5 je histogram rozdělení četností, který na vodorovné ose zachycuje intervaly, do kterých jsou data rozdělena pomocí cenzoru a na svislé ose vidíme počet výskytu těchto dat v daném intervalu, což v našem případě je zároveň očekávaný reálný počet zákazníků. Na obrázku 3.5 sloupec 7 + představuje všechny četnosti sloučené dohromady pomocí cenzoru.

Tabulka 3.7 Data pro sestavení histogramu četností pro trénovací období

Třída	0	1	2	3	4	5	6	7+
Skutečná	796	284	177	93	68	45	33	168
Očekávaná	790,9	292,4	162,6	104,3	72,1	52,2	38,9	150,6

Zdroj: vlastní výpočet v R

Na závěr k tomuto grafu můžeme říct, že náš model generuje očekávané výsledky pomocí reálných dat v trénovacím období poměrně dobře. Dále zkusíme ověřit, že model udržuje svou stabilitu a kvalitu i ve validačním období. Zprv je třeba najít informace o validačním období. Pomoci speciálního příkazu v R vytvoříme jednu společnou matici, která bude zahrnovat trénovací i validační období. Tato struktura představuje kombinaci dvou období za účelem zjištění počtů transakcí jednotlivých klientů, které byly realizované ve validačním období. Nicméně, samotnou informaci získáme přímo ze základního souboru – předem odečteme opakované objednávky v trénovacím období z celkového počtu transakcí. Tento krok provádíme až tehdy, když se nechceme soustředit právě na počet opakovaných objednávek.

Stejně jako jsme posuzovali kvalitu trénovacího období, nyní můžeme porovnat skutečná a očekávaná data ve validačním období.



Obrázek 3.6 Frekvence opakovaných objednávek ve validačním období

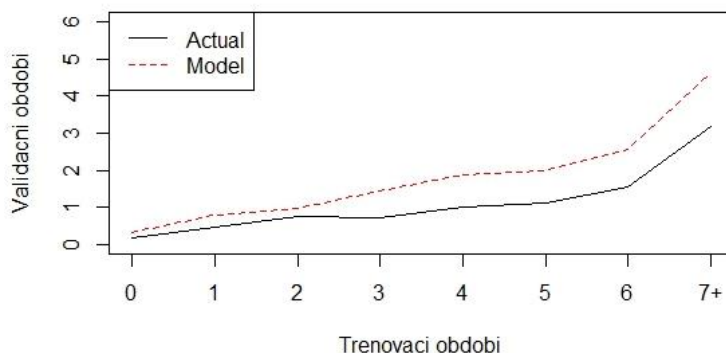
Tabulka 3.8 obsahuje použitá data pro konstrukci histogramu četnosti, odpovídající validačnímu období (viz graf 3.6). Z grafu 3.6 vidíme podobné hodnoty mezi skutečnými a očekávanými daty a tím pádem jsme ověřili, že model udržuje svou stabilitu a kvalitu i ve validačním období.

Tabulka 3.8 Data pro sestavení histogramu četností pro validační období

Třída	0	1	2	3	4	5	6	7+
Skutečná	392	130	65	36	28	14	9	26
Očekávaná	384,7	135,1	68,1	39,1	24	15,3	10	23,8

Zdroj: vlastní výpočet v R

Ted' můžeme graficky zobrazit vztah mezi validačním a trénovacím obdobím. Obrázek 3.7 porovnává trénovací a validační období, což poskytuje dostatečný přehled o zákaznickém chování. Na daném grafu také vidíme, že zákazníci jsou roztrženi do stejného počtu skupin pomocí cenzoru 7.



Obrázek 3.7 Vztah mezi trénovacím a validačním obdobím

Pro lepší znázornění a přehlednost rozdílů mezi kvalitou predikcí dat ve validačním a trénovacím obdobím slouží následující tabulka, která zahrnuje rozdíly mezi skutečnými a očekávanými daty pro obě období.

Tabulka 3.9 Srovnání odchylek skutečných a očekávaných dat

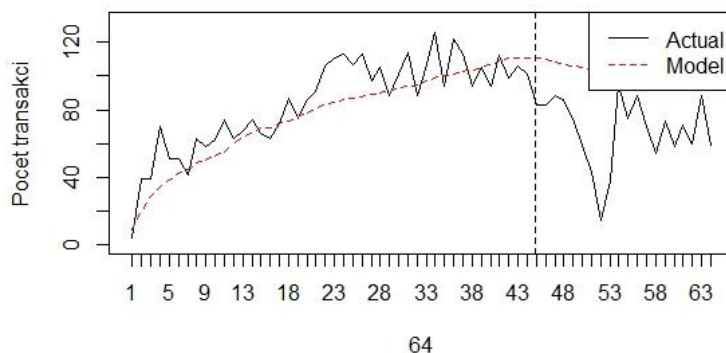
Třída	0	1	2	3	4	5	6	7+
Trénovací	5,1	-8,4	14,4	-11,3	-4,1	-7,2	-5,9	17,4
Validační	7,3	-5,1	-3,1	-3,1	4	-1,3	-1	2,2

Zdroj: vlastní výpočet v R

Tabulka 3.9 obsahuje odchylky očekávaných dat od skutečných, jak pro trénovací období, tak i pro validační. Občas je obtížné rozdíly mezi těmito ukazateli „vyčíst“ z grafů (obrázky 3.5 a 3.6), v těchto případech sledujeme rozdíl mezi černými a červenými sloupci našich histogramů četností. Z tabulky 3.9 můžeme zjistit, že rozdíly v prvních třídách jsou především u trénovacího období menší než v ostatních skupinách, u kterých je rozdělen datový soubor pomocí cenzoru. Celkové odchylky od skutečných dat se pohybují v rozmezí od -11,3 do 17,4, což je poměrně uspokojivé, protože většina odchylek má mnohem menší hodnoty než krajní hodnoty intervalu.

Seskupení podle četností ve validačním období je jedním ze způsobů popisu zákaznického chování. V následujícím kroku zkusíme roztrždit objednávky podle času: ověříme, jak přesně

náš model predikuje počet transakcí, který se objeví každý týden. Použijeme klasickou datovou matici z našeho modelu, která již obsahuje údaje seskupené podle časových úseků. Využijeme tuto skutečnost pro zjištění celkového počtu objednávek za určitý den. Dále převedeme denní frekvenci na týdenní pro lepší orientaci ve výsledcích.

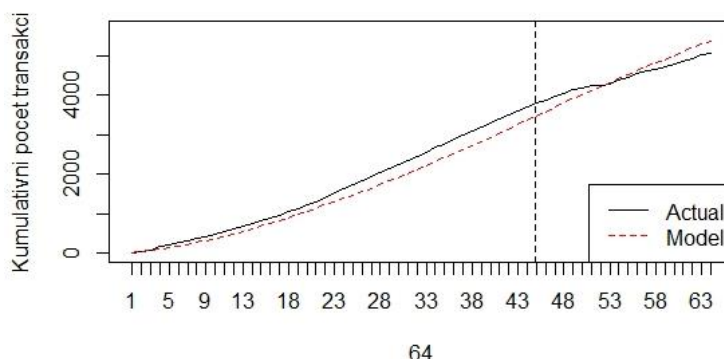


Obrázek 3.8 Srovnání skutečného a očekávaného počtu transakcí za celkové období

Obrázek 3.8 zobrazuje skutečný počet transakcí v porovnání s namodelovaným počtem za celé sledované období (v našem případě 64 týdnů). Na ose Y je samotný počet vyskytujících se transakcí a osa X představuje jednotlivé týdny. Skutečná datová řada, která je v grafu zobrazena černou čarou, je velice dynamická, proto ji vyhlazujeme pomocí datové řady modelu, která zachycuje hlavní trend vývoje transakcí v čase. Největší výskyt transakcí pozorujeme přibližně v průběhu 35. týdne. Významný propad objednávek pozorujeme po padesátém týdnu. Toto období odpovídá od prosince roku 2015 až do prvního týdne ledna 2016. Výrazný pokles je způsoben především obdobím vánočních svátků. Takový trend lze pozorovat jak na trzích časopisů a novin, tak i na trhu práce. Model Pareto/NBD nepočítá se sezonními výkyvy, a proto nadhodnocuje očekávaný počet objednávek v průběhu týdnů 50 až 53. Skutečná datová řada je velmi různorodá v závislosti na týdnu, a proto jsme konstruovali adaptivní řadu, která modeluje počet transakcí. Pomocí tohoto grafu máme možnost predikovat zákaznické chování v následujících obdobích a připravit vhodné strategie pro každý případ.

Následující graf (3.9) představuje skutečný a očekávaný celkový počet opakujících se objednávek za celé trénovací i validační období pro všechny zákazníky. Na tomto grafu jsme zkusili vyhladit datovou řadu pomocí kumulativního zachycení objednávek, takže z rostoucího trendu opravdu vidíme kumulativní nárůst počtu objednávek s časem. Zajímavé je také to, že na začátku sledovaného období model Pareto/NBD popisuje zákaznické chování mnohem

přesněji než na konci tohoto období. Ale celkem vybraný model funguje správně a vypadá důvěryhodně.



Obrázek 3.9 Srovnání skutečného a očekávaného kumulativního počtu transakcí

V následující subkapitole zhodnocujeme výpočty a uvedeme závěry, plynoucí z grafických a matematických výstupů. Také zkusíme vypočítat celoživotní hodnotu podle teoretických předpokladů, uvedených v podkapitole 2.4, která se věnovala popisu fungování modelu Pareto/NBD.

3.5 Zhodnocení modelu Pareto/NBD

Některé obecné závěry o modelu můžeme odvodit i z výše uvedených grafů, ale pro ověření, zda model Pareto/NBD popisuje data ve validačním a trénovacím období dostatečně kvalitně, nám nestačí pouze grafický výstup; je třeba vypočítat základní statistiku RMSE (Root Mean Squared Error), na jejímž základě posoudíme kvalitu modelování. RMSE je střední odmocninová chyba, kterou můžeme spočítat dle následujícího vzorce:

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2} \quad (3.2),$$

kde n je počet pozorování,

y_j – skutečné hodnoty,

\hat{y}_j – odhadované hodnoty.

Daná statistika slouží především pro vzájemné porovnání predikovaných a skutečných hodnot, sama o sobě je těžko interpretovatelná, a z názvu je patrné, že čím statistika je menší, tím je model kvalitnější (Armstrong a Collopy, 1992). Z tohoto důvodu uvedeme několik

ukazatelů RMSE pro různé případy. Výsledky zobrazíme ve formě tabulky 3.10 a následně okomentujeme.

Tabulka 3.10 Zhodnocení trénovacího období

Úroveň	RMSE	RMSE, %
Zákazník	0,233	0,637
Týden	0,174	0,451

Zdroj: vlastní výpočet v R

Nejprve zhodnotíme pomocí střední odmocninové chyby trénovací období. Tabulka 3.10 obsahuje dvě základní statistiky RMSE: první řádek popisuje kvalitu predikce na úrovni zákazníka a druhý řádek na týdenní úrovni. Pro určení první odchylky (0,233) jsme spočítali rozdíl skutečného a predikovaného počtu transakcí za celé trénovací období a pak tyto hodnoty zprůměrovali, čímž jsme vlastně získali dobrý základ pro vyhodnocení kvality modelování v trénovacím období. Tato metrika nám ukazuje, jak je model dobrý ve vybírání cílových zákazníků, na které bychom se měli zaměřit v rámci marketingu. Dále uvádíme hodnoty RMSE na úrovni týdnů, kde porovnáme skutečný a očekávaný počet transakcí všech zákazníků v daném týdnu. Spočítali jsme rozdíl mezi těmito hodnotami a pak stanovili průměr přes všechny týdny za celé trénovací období. Daný ukazatel nám říká, jak se na model můžeme spolehnout z hlediska predikce celkového počtu transakcí (pro lepší plánování zdrojů a ostatních složek organizační strategie). Hodnota střední odmocninové chyby za celé období je 0,174, což v porovnání se zákaznickou hodnotou nám umožňuje říct, že model v trénovacím období popisuje data stále dostatečně kvalitně. Tabulka 3.10 obsahuje i takzvanou normalizovanou hodnotu RMSE v procentech, která byla spočítaná podle následujícího vzorce:

$$\%RMSE = \frac{RMSE}{\frac{\sum_{i=1}^n x_i}{n}} * 100 \quad (3.3),$$

kde ve jmenovateli je aritmetický průměr pozorované proměnné.

První hodnota RMSE v procentech z průměrného počtu transakcí je spočítaná z průměru na zákazníka (0,637%), druhá z průměru na týden (0,451%). Hodnota odmocninové chyby v procentech nám ukazuje, jak moc se mýlíme ohledně predikování hodnot ve validačním období, což je patrné z názvu statistiky. Týdenní predikce, která v budoucnu slouží k podpoře strategického rozhodování, má relativně lepší hodnoty než predikce na úrovni zákazníka.

Ted' provedeme výpočty i pro validační období, abychom měli komplexní základ pro porovnání a vyhodnocení modelu Pareto/NBD.

Tabulka 3.11 Zhodnocení validačního období

Úroveň	RMSE	RMSE, %
Zákazník	0,853	2,113
Týden	0,777	1,676

Zdroj: vlastní výpočet v R

Podle stejného algoritmu porovnáme rozdíl mezi predikovanými a skutečnými hodnotami za celé validační období. Na základě údajů uvedených v tabulce 3.11 můžeme říct, že RMSE v trénovacím období nám vyšlo mnohem lépe ve srovnání s validačním, a to jak na zákaznické úrovni, tak i pro týdenní frekvenci transakcí. Jak už jsme zmínili, neexistuje přesné stanovení škály „dobrých“ a „špatných“ hodnot statistiky RMSE. Vzhledem k našemu datovému souboru (rozsah dat je relativně velký v porovnání s výsledkem RMSE) můžeme odvodit, že hodnoty odchylek/chyb se pohybují v přijatelném rozmezí. Všechny výpočty uvedené v tabulkách 3.10 a 3.11 nyní propojíme s obrázkem 3.8 a vidíme, že hodnoty RMSE jasně navazují na grafické výstupy. Podle tohoto obrázku je evidentní, že modelování očekávaných hodnot v trénovacím období vychází kvalitněji než v období validačním, kde lze pozorovat mnohem větší variabilitu hodnot a odlišnost očekávaných hodnot od skutečných (červená a černá čára na obrázku 3.8). Výsledky srovnání predikční schopnosti a posouzení kvality modelu Pareto/NBD potvrzujeme i pomocí statistiky RMSE (tabulky 3.10 a 3.11).

3.6 Finanční analýza efektivity modelu Pareto/NBD

V této podkapitole je hlavním cílem popsat použitelnost výstupů modelování v praxi a ukázat na reálných datech přínos pro firmu z udržení cílových zákazníků. Jak už bylo zmíněno v teoretické kapitole, nemusíme si držet libovolného klienta za každou cenu, ale vybereme nejziskovější segmenty zákazníků a zaměříme se především na ně. Z výstupu modelu Pareto/NBD určíme též pravděpodobnost, že zákazník je stále aktivní na konci trénovacího období a také známe očekávaný počet transakcí v období validačním. Na základě těchto údajů je možné vybrat klienty s největší celoživotní hodnotou a přizpůsobit jim odpovídající marketingové nástroje, jako je například poskytnutí slevy. Ze strany firmy je poskytnutí slevy vždy nákladem, výnos klesá stejně, jako se snižuje zisk společnosti. Ale současně pomocí racionální slevové politiky zvyšujeme pravděpodobnost, že klient bude nakupovat stále u nás a neodejde ke konkurenci. V návaznosti na důležitost správného rozhodnutí při výběru cílových zákazníků a výše slevy použijeme výpočet CLV k individualizaci výše poskytnuté slevy jednotlivým zákazníkům. Jak už víme z výstupů modelu Pareto/NBD, tak naši zákazníci jsou velmi odlišní: mají různé nákupní chování a velkou variabilitu sumy transakcí a frekvence

objednávek, a proto dávat slevu plošně je zbytečně nákladné a neefektivní. Náklady sice rostou při poskytnutí slevy, ale tímto způsobem zabráníme odchodu některých klientů a předpokládáme, že při správném výběru cílových zákazníků obdržíme větší zisk. V průběhu této kapitoly spočítáme ukazatel ROI (Return On Investments), který stanovuje návratnost investic jako poměr příjmů z investic a nákladů pomocí CLV (Licino, 2013):

$$ROI, \% = \frac{CLV - Investice}{Investice} * 100 \quad (3.4).$$

V prvním kroku byla provedena ABC analýza, kde je cílem setřídít klienty podle ziskovosti za celkové období. V našem případě má firma velký počet stejně ziskových zákazníků, a proto nemůžeme vybrat určité „topové“ klienty. Řešíme to pomocí ABC skupin, kde každá skupina je reprezentativní a závěry o ní se dají vztáhnout na jednotlivé klienty, kteří skupinu tvoří. Každá třída odpovídá určitému segmentu podle sumy objednávek.

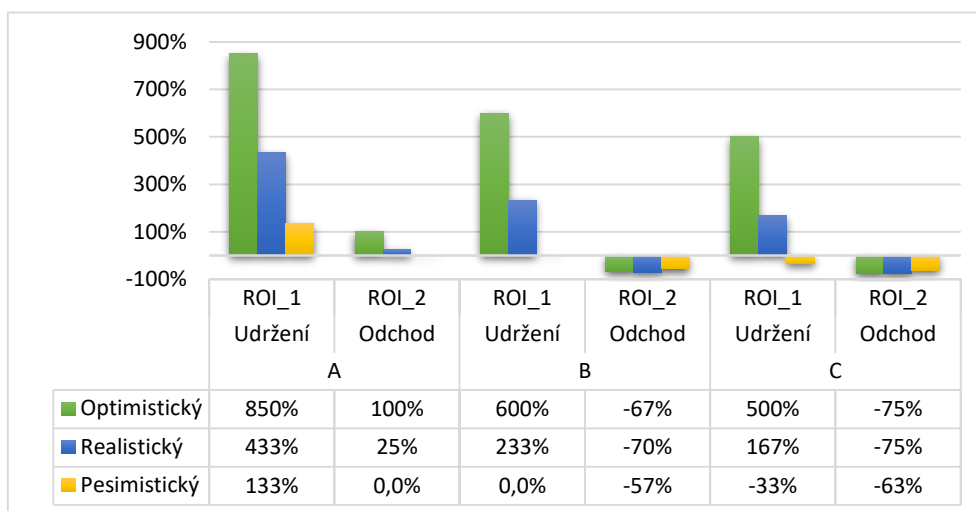
Tabulka 3.12 Fragment průběžné přípravy datového souboru

	Počet klientů	Podíl počtu	Suma objednávek, Kč	Podíl sumy	Průměrná objednávka za týden, Kč
A	470	22%	5 991 825	79%	698
B	858	40%	1 331 749	18%	318
C	810	38%	232 280	3%	152
Celkem	2 138	100%	7 555 854	100%	X

Zdroj: vlastní výpočet v Excelu

V tabulce 3.12 jsou uvedeny výstupy roztržidění všech zákazníků do tří ABC skupin. Zde je vidět, že skupina A, která obsahuje 22 % všech klientů, vytváří téměř 80 % zisku. Což je velký rozdíl ve srovnání se skupinou C, ve které je 38 % klientů a tvoří pouze 3 % z celkového zisku firmy za sledované období. Dále byla provedena finanční analýza náhodných klientů z každé skupiny takovým způsobem, že jsme prokázali důležitost přizpůsobení slevové politiky dle cílových zákazníků (viz přílohy 1,2,3). Pomocí celoživotní hodnoty klienta a pravděpodobnosti aktivity na konci trénovacího období podle všech tří scénářů vývoje spočítáme výnosy z udržení jednotlivých klientů a poté stanovíme tři varianty velikosti slev, které považujeme za náklady a očekávaný počet objednávek. Optimistická prognóza počítá se slevou 10 %, realistická 15 % a pesimistická 30 %. Všechny údaje potřebné pro tento výpočet byly získány z modelu Pareto/NBD v programu R. Nakonec podle vzorce 3.4 stanovíme velikost návratnosti investic v procentech, kde za investice považujeme velikost poskytovaných slev a za výnosy hodnotu CLV násobenou pravděpodobností aktivity klienta. Celoživotní hodnotu jsme

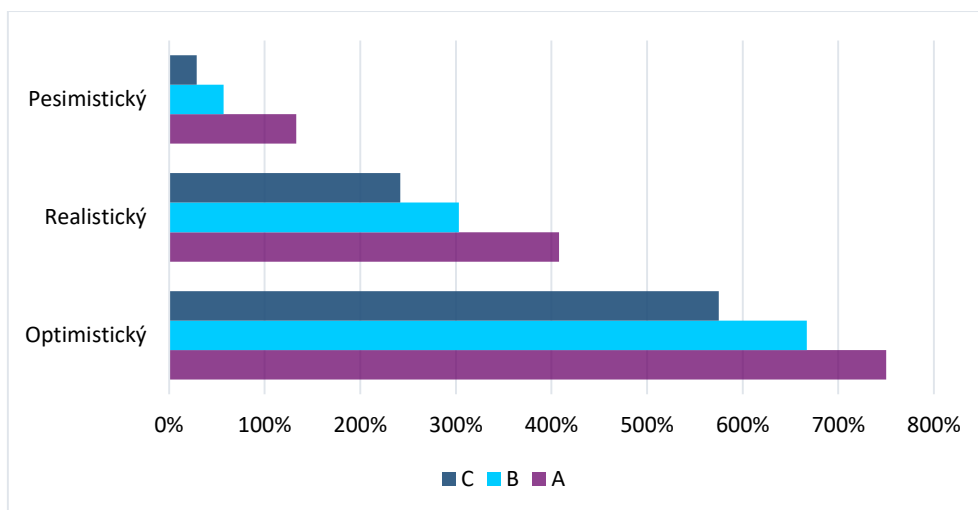
spočítali pro každou skupinu zvlášť pomocí průměrné týdenní objednávky z tabulky 3.12. Přílohy 1, 2 a 3 obsahují popsany výpočet ROI pro zákaznické skupiny A, B, C.



Obrázek 3.10 Návratnost investic dle různých scénářů

Zdroj: vlastní zpracování v Excelu

Obrázek 3.10 nám ukazuje hodnoty ROI podle tří scénářů vývoje v případě udržení klienta (poskytnutí slevy) a v opačném případě (klient by odešel ke konkurenci). Vidíme, že u všech skupin existuje značný rozdíl mezi hodnotami ROI a dopadem z poskytování slev. Ale nejvýnosnější segment patří klientům skupiny A podle optimistické varianty vývoje (tím myslíme slevu 10 %, pravděpodobnost aktivity na konci trénovacího období 0,95 a očekávaný počet objednávek 3,7). Z výstupu R můžeme stanovit tyto údaje a pokusit se podle nich změnit velikost slevy; tím stanovíme nejziskovějšího klienta a jeho optimální slevu. Například pro skupinu B maximální sleva činí 15 %, jinak je hodnota ROI rovna nule (příloha 2). Skupině A firma může poskytnout i větší slevu, každopádně i podle pesimistické prognózy návratnost investic je vyšší než 100%. Skupinu C také nemá smysl si držet za každou cenu (z komerčního hlediska), větší sleva má za následek pouze finanční ztrátu. Druhý sloupec každé skupiny obsahuje hodnotu ROI v případě odchodu klienta, kdybychom neposkytnuli slevu, a s určitou pravděpodobností je stanoven ušlý zisk. Následující graf 3.11 zobrazuje přesně ten rozdíl mezi ROI z udržení klienta a ROI v případě, že klient odešel ke konkurenci nebo došlo k jeho likvidaci. Čím větší je rozdíl mezi těmito ukazateli, tím výhodnější je investovat peníze do zákaznického segmentu a poskytovat slevu jednotlivým zákazníkům.



Obrázek 3.11 Rozdíly mezi ROI_1 a ROI_2

Zdroj: vlastní zpracování v Excelu

Například vybereme-li klienta ze skupiny A, předpokládáme maximální slevu v 30 % podle pesimistického scénáře: z grafu 3.11 je vidět, že i v tomto případě rozdíl v příjmech stále existuje. Tím pádem můžeme říct, že když zabráníme odchodu klienta ze skupiny A pomocí maximální slevy, tak se budeme nacházet v lepším postavení, než když povolíme klientovi od firmy odejít. Skupina C nemá až tak značný rozdíl a zjevnou výhodu z hlediska návratnosti investic. Zde zatím není důvod poskytovat velkou slevu. Ale v rámci optimistické varianty a u slevy 10 % už je zřejmá výhoda, protože oproti velmi malým nákladům (10 % z průměrné objednávky skupiny C v tabulce 3.12) stojí několik drobných nákupů od této skupiny, které jsou z celkového pohledu ziskovější a přesahují mnohonásobně poskytnutou slevu.

Konstruované výpočty na základě modelu Pareto/NBD umožňují firmě simulovat různé scénáře podle velikosti slev, pravděpodobnosti aktivity klienta a očekávaného počtu transakcí v budoucnu. V byznysu se u malých a středních podniků často stává, že firma má 10 až 15 velkých zákazníků, které si drží. Z tohoto důvodu si nemůžou dovolit naslepo odhadovat velikost slevy pro každého zákazníka. Je nutné, aby každé marketingové rozhodnutí mělo kvantitativní opodstatněnost. Chybné řešení má v tomto případě příliš vysokou cenu. Pomocí výstupů modelu Pareto/NBD se dá kvalitně vybrat cílových zákazníků a vědět dopředu, kolik investujeme peněz a kolik se nám vrátí zpátky. V současné realitě konkurenčních tlaků jsou analýza výnosnosti a predikce očekávané návratnosti investic nedílnými součástmi podnikových strategií. Dalším námětem pro zkoumání a možným rozšířením CLV analýzy by bylo na ni navázat s modely odchodovosti zákazníků (churn management) a mírou odchodu (churn rate). Míra odchodu může být důležitým vstupem do modelování CLV a simulace ROI. Díky těmto přístupům by mohli manažeři firmy na základě zákaznického chování odhadnout,

kteří z klientů společnosti pravděpodobně zvažují přechod ke konkurenci. Pomoci analýzy celoživotní hodnoty je možné vyhodnotit, jestli má smysl do těchto zákazníků investovat a dávat jim retenční nabídku (slevy či jiné benefity) nebo jestli odejdou za všech okolností.

Závěr

Bakalářská práce se zabývala aplikací statistických metod při výpočtu celoživotní hodnoty zákazníka, popisem různých přístupů k modelování CLV a řízením vztahů se zákazníky pomocí modelu Pareto/NBD. Tato práce patří mezi empirické typy prací, kde teoretické předpoklady byly otestovány v praxi na reálných datech firmy „Práce Pro Vás“. V průběhu práce bylo prozkoumáno klientské chování s cílem přizpůsobení marketingové strategie, výběru cílových zákazníků a efektivní alokaci zdrojů. Hlavní kvantitativní výsledky modelu Pareto/NBD byly propojeny s grafickými výstupy pro lepší znázornění a vizualizaci.

Na úvod bylo vymezeno téma a důvod pro potřebu zkoumání celoživotní hodnoty zákazníka pomocí statistických metod. Klíčová motivace spočívala v potřebě kvantifikovat výstupy marketingových aktivit a podpořit strategické rozhodování.

V první teoretické kapitole jsme popsali koncept CLV a různé přístupy k modelování zákaznického chování na základě minulých dat. Zjistili jsme, že hodnota konkrétního klienta je velmi důležitá pro zjištění nejziskovějších vztahů společnosti z komerčního hlediska. Každý podnik při strategickém plánování očekává určitý zisk a přínos své činnosti. Celoživotní hodnota zákazníka představuje ten samý budoucí výnos pro firmu. V průběhu kapitoly jsme charakterizovali šest různých modelů a přístupů, které se používají k odhadu CLV: RFM modely, ekonometrické modely, modely založené na analýze časových řad, modely strojového učení, modely růstu a pravděpodobnostní modely. Vybrali jsme konkrétní pravděpodobnostní model Pareto/NBD pro aplikaci na reálných datech společnosti „Práce Pro Vás“, který byl velmi podrobně popsán v subkapitole 2.4 včetně předpokladů, parametrů a odvození vzorců pomocí jednotlivých pravděpodobnostních rozdělení.

Druhá kapitola zahrnuje úplný přehled a popis všech populárních přístupů v oblasti zkoumání a predikce celoživotní hodnoty zákazníka, významných šetření v této oblasti a hlavního přínosu jednotlivých šetření od roku 1994. Zmínili jsme pojem CRM, jeho účel při určení dlouhodobých vzájemných kontaktů a potvrdili jsme skutečnost, že je velmi důležité se v současné době věnovat zákazníkům, diferencovat a segmentovat přístupy k nim a také pravidelně vyhodnocovat, kontrolovat a korigovat firemní chování a péči o zákazníka.

V další části byly splněny hlavní praktické cíle bakalářské práce. Klíčovým záměrem v průběhu třetí kapitoly práce bylo aplikovat teoretický model Pareto/NBD na empirický příklad společnosti „Práce pro Vás“ a predikovat celoživotní hodnotu zákazníků pomocí programovacího jazyka R. K tomu jsme dospěli pomocí předběžné přípravy dat, která je velmi

důležitou fází každé analýzy. V případě vyhodnocení kvality modelování byly spočítány statistiky RMSE pro různé případy.

Za hlavní poznatky a výsledky z modelování považujeme podkapitolu 3.6, kde byla provedena finanční analýza efektivity modelu Pareto/NBD pomocí ukazatelů návratnosti investic. Na základě výstupů z uvedeného modelu je možné vybrat klienty s největší celoživotní hodnotou a přizpůsobit odpovídající marketingové nástroje, například poskytnutí slevy. Ze strany firmy poskytnutí slevy je vždy snížení výnosů, výnos klesá stejně, jako se snižuje i zisk společnosti. Ale současně pomocí racionální slevové politiky zvyšujeme pravděpodobnost, že klient bude nakupovat stále u nás a neodejde ke konkurenci. V návaznosti na důležitost správného rozhodnutí při výběru cílových zákazníků a výše slevy použijeme výpočet CLV k individualizaci výši poskytnuté slevy jednotlivým zákazníkům.

Za základní přínos této bakalářské práce k řešené problematice považuji teoretické zpracování skutečných přístupů k modelování celoživotní hodnoty zákazníka a následné praktické využití výstupů modelu Pareto/NBD pro dlouhodobý záměr firmy. Výsledky aplikace uvedeného pravděpodobnostního modelu společnost může efektivně využívat ve své každodenní praxi a tím získávat kvantitativní podstatu pro vybírání cílových zákazníků a zaměření svých marketingových úsilí. K hlavnímu přínosu pro firmy patří i zmíněná finanční analýza efektivity na základě výstupů z modelu Pareto/NBD, kde byly probrány tři scénáře nákupního chování zákazníků ovlivněných jednotlivými slevovými politikami, což umožňuje firmě podporovat strategická rozhodování a individualizovat výši poskytnuté slevy jednotlivým zákazníkům.

Mezi další náměty pro zkoumání v této oblasti patří aplikace jiných statistických modelů popsanych v teoretické části na reálná data a porovnání výsledků modelování mezi sebou. Také zkoumání může být pokračováno využitím a zhodnocením delší časové řady (nyní sledujeme časový úsek dlouhý o 64 týdnů) a zhodnocení výstupů Pareto/NBD na delším období. Další možností zkoumání a možným rozšířením CLV analýzy by bylo doplnění o modely odchodovosti zákazníků (churn management) a mírou odchodu (churn rate). Míra odchodu může sloužit jako důležitý vstup do modelování CLV a simulace ROI.

Na začátku jsme stanovili hlavní cíle práce:

- ✓ popsat různé přístupy k modelování CLV;
- ✓ aplikovat model Pareto/NBD na reálná data společnosti „Práce pro Vás“;
- ✓ zhodnotit využitelnost výsledků predikování CLV v praxi.

V teoretické kapitole byly popsány jednotlivé přístupy a metody výpočtu celoživotní hodnoty, také stručně byla představena rešerše literatury a přehled šetření v oblasti modelování

CLV od roku 1994 až do roku 2011. Praktická část obsahuje aplikaci modelu Pareto/NBD na data firmy, zhodnocení výsledků modelování. V závěru této kapitoly byla uskutečněna finanční analýza návratnosti investic a dopady různých přístupů slevové politiky společnosti. Všechny cíle se podařilo splnit, čímž jsme dosáhli příznivých výsledků a přispěli k zefektivnění marketingové činnosti firmy.

Seznam literatury

ARMSTRONG J. S., COLLOPY F. (1992). Error Measures for Generalizing About Forecasting Methods: Empirical Comparisons. *International Journal of Forecasting*, č. 8, s. 69–80. URL:

http://repository.upenn.edu/cgi/viewcontent.cgi?article=1075&context=marketing_papers.

BARAN R., GALKA J. (2013). *CRM: the foundation of contemporary marketing strategy*. New York, NY: Routledge.

DEKIMPE M. G., HANSENS D. M. (2005). Persistence Models and Marketing Strategy. *Tijdschrift voor Economie en Management*, č. 5, s. 69–93.

DONKERS B., VERHOEF P. C., MARTIJN G. DE JONG (2007). Modeling CLV: A test of competing models in the insurance industry [online]. *Springer Science + Business Media*. URL: <http://link.springer.com/article/10.1007/s11129-006-9016-ya>.

DWYER F. R., SCHURR P., Oh S. (1987). Developing Buyer-Seller Relationships. American Marketing Association: *Journal of Marketing*, č. 51, s. 11–27.

ESMAEILIGOOKEH M., TAROKH M. J. (2013). Customer Lifetime Value Models: A literature Survey. *International Journal of Industrial Engineering & Production Research*, č. 24, s. 317–336. URL: <http://ijiepr.iust.ac.ir/article-1-509-en.pdf>.

FADER P. S., HARDIE B. (2005). *A Note on Deriving the Pareto/NBD Model and Related Expressions*. URL: <http://www.brucehardie.com/notes/009/>.

FADER S. P., HARDIE B. (2009). Probability Models for Customer-Base Analysis. *Journal of Interactive Marketing*, č. 23, s. 61–69. URL: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.180.2024&rep=rep1&type=pdf>.

FADER P. S., HARDIE B., LEE K. L. (2005). Counting Your Customers the Easy Way: An Alternative to the Pareto/NBD Model. *Marketing Science*, č. 24, s. 275–284. URL: <http://pubsonline.informs.org/doi/abs/10.1287/mksc.1040.0098?journalCode=mksc>.

FADER P. S., HARDIE B., LEE K. L. (2005). *RFM and CLV: Using Iso-value Curves for Customer Base Analysis*.

FADER P. S., HARDIE B., SHANG J. (2010). Customer-Base Analysis in a Discrete-Time Noncontractual Setting. *Marketing Science*, č. 29, s. 1086–1108.

FLORDAL P., FRIBERG J. (2013). *Modeling Customer Lifetime Value in the Telecom Industry*. Lund University, Stockholm. URL: http://www.pm.lth.se/fileadmin/pm/Exjobb/Exjobb_2013/Flordahl___Friberg/CLV_ERICSSON_Flordal_Friberg.pdf

FRIEDMAN J. H. (1991). Multivariate Adaptive Regression Splines. Institute of Mathematical Statistics. *The Annals of Statistics*, č. 19, s. 1–67.

- GLADY N., BAESENS B., CROUX C. (2009). *A Modified Pareto/NBD Approach for Predicting Customer Lifetime Value*. Catholic University of Leuven.
- GUPTA S., a kol. (2006). Modeling Customer Lifetime Value. *Journal of Service Research*, č. 9, s. 139–155. URL: <http://jsr.sagepub.com/content/9/2/139.abstract>.
- GUPTA S., DONALD R. LEHMANN (2010). *Managing Customers as Investments*. 1. vyd. In: Wharton School Pub.
- HASTIE T., TIBSHIRANI R. (1990). *Generalized additive models*. New York: Chapman and Hall.
- HOGAN, JOHN E., LEMON N., LIBAI B. (2003). Quantifying the Ripple: Word-of-Mouth and Advertising Effectiveness. *Journal of Advertising Research*.
- JANGID C., KOTHARI T. (2013–2014). *CustoVal: Estimating Customer Lifetime Value Using Machine Learning Techniques*.
- JERATH K., FADER P. S., HARDIE B. (2011). New Perspectives on Customer “Death” Using a Generalization of the Pareto/NBD Model. *Marketing Science*, č. 30, s. 866–880. URL: https://www0.gsb.columbia.edu/mygsb/faculty/research/pubfiles/6057/customer_death.pdf.
- KARLÍČEK M., KRÁL P. (2011). *Marketingová komunikace: jak komunikovat na našem trhu*. Praha: Grada.
- KOŘENÁŘ V. (2002). *Stochastické procesy*. Praha: Vysoká škola ekonomická.
- KOTLER P. (2000). *Marketing podle Kotlera: jak vytvářet a ovládnout nové trhy*. Praha: Management Press.
- KUMAR V., LUO, MAN A. (2006). Linking an Individual's Brand Value to The Customer Lifetime Value: An Integrated Framework. In: *AMA Winter Educators' Conference Proceedings*, č. 17, s. 152.
- KUMAR V. (2008). *Customer Lifetime Value the Path to Profitability*. Boston: now Publishers.
- KUMAR V., GEORGE M. (2007). Measuring and maximizing customer equity: a critical analysis. *Journal of the Academy of Marketing Science*, č. 35, p.157–171.
- KUMAR, V., REINARTZ W. (2012). *Customer Relationship Management Concept, strategy, And Tools*. 2 vyd. Berlin: Springer.
- KUMAR V., SHAH D. (2015). *Handbook of research on customer equity in marketing*. Cheltenham: Edward Elgar.
- LEHTINEN J. (2007). *Aktivní CRM: řízení vztahů se zákazníky*. Praha: Grada.
- MALÁ, I. (2013). *Statistické úsudky*. 1. vyd. Praha: Professional Publishing.
- MAREK L. (2012). *Pravděpodobnost*. Praha: Professional Publishing.

McCARTHY D. a kol. (2014). Implementing Buy 'Til You Die Models [online]. URL: <https://cran.r-project.org/web/packages/BTYD/BTYD.pdf>

McCARTHY D., WADSWORTH E., a kol. (2014). Buy 'Til You Die - A Walkthrough [online]. URL: <https://cran.r-project.org/web/packages/BTYD/vignettes/BTYD-walkthrough.pdf>

REICHHELD, FREDERICK F. a TEAL T. (2001). *The loyalty effect: the hidden force behind growth, profits, and lasting value*. Boston, Mass.: Harvard Business School Press.

REINARTZ W., KRAFFT M., HOYER W. (2004). The Customer Relationship Management Process: Its Measurement and Impact on Performance. *Journal of Marketing Research*, č. 41, s. 293–305.

REINARTZ W., KUMAR V. (2004). *Getting the Most Out of All Your Customers*. Harvard Business Review, č. 82, p. 116–23.

REINARTZ W., THOMAS J., KUMAR V. (2005). Balancing Acquisition and Retention Resources to Maximize Customer Profitability. *Journal of Marketing*, č. 69, s. 63–79.

RUST a kol. (2004). Return on Marketing: Using Customer Equity to Focus Marketing Strategy. *Journal of Marketing*, č. 68, s. 109–127.

RUST, LEMON N, ZEITHAML A. (2004). Return on Marketing: Using Customer Equity to Focus Marketing Strategy. *Journal of Marketing*, č. 68, s. 109–127.

SCHMITT B. (c2003). *Customer experience management: a revolutionary approach to connecting with your customers*. New York: Wiley.

SCHMITTLEIN, MORRISON G., COLOMBO R. (1987). *Counting your customers: who are they and what will they do next?* Management Science.

SHABANOVA L. B., a kol. (2015). ABC - Analysis, as an Important Tool for Generating an Optimal Assortment Plan Commercial Enterprises. *Mediterranean Journal of Social Sciences*, CSER Publishing. Řím, Italie, č. 6, s. 691–694.

SOEINI R. A., FATHALIZADE E. (2012). Customer Segmentation based on Modified RFM Model in the Insurance Industry. In: *4th International Conference on Machine Learning and Computing. IPCSIT*, č. 25 IACSIT Press, Singapore. URL: http://www.icmlc.org/icmlc2012/020_icmlc2012.pdf.

SOHRABI S., NIKKHAHAN B. (2011). Value Creation for Customers by Segmentation Approach: An Online Game and Entertainment Website Case Study. *International Journal of Digital Content Technology and its Applications*, č. 5.

SRINIVASAN S., HANSSENS D. M. (2007). Marketing and Firm Value. Metrics, Methods, Findings and Future Directions. URL: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.146.470&rep=rep1&type=pdf>.

TAVANA A. F., a kol. (2013). Theoretical Models of Customer Relationship Management in Organizations. *International Journal of Business and Behavioral Sciences*, č. 3, s. 63–70.

VENKATESAN R., KUMAR V. (2004). A Customer Lifetime Value Framework for Customer Selection and Resource Allocation Strategy. *Journal of Marketing*, č. 68, s. 106–125.

VILLANUEVA a kol. (2008). The Impact of Marketing-Induced Versus Word-of-Mouth Customer Acquisition on Customer Equity Growth. *Journal of Marketing Research*, č. 45, s. 48–59.

VRANÁ L., JAŠEK P. (2014). Managerial Impacts of Different Computation Models for Customer Lifetime Value for an E-Commerce Company. In: *8th International Days of Statistics and Economics*. Vysoká škola ekonomická, Praha. URL: https://msed.vse.cz/msed_2014/article/333-Jasek-Pavel-paper.pdf.

VRANÁ L., JAŠEK P. (2015). Persistence Models for Customer Equity. In: *9th International Days of Statistics and Economics*. Vysoká škola ekonomická, Praha. URL: https://msed.vse.cz/msed_2015/article/85-Vrana-Lenka-paper.pdf.

WITTEN I. H., FRANK, HALL A. (c2011). *Data mining: practical machine learning tools and techniques*. 3. vyd. Burlington: Morgan Kaufmann series in data management systems.

Seznám obrázků a tabulek

Obrázek 2.1 Koncept modelování CLV.....	4
Obrázek 2.2 Vývoj CRM systému.....	18
Obrázek 2.3 Agregace kumulativního počtu transakcí.....	23
Obrázek 2.4 Příklad transakčního chování jednotlivých zákazníků	23
Obrázek 2.5 Proces modelování a predikce transakčního chování	24
Obrázek 3.1 Průměrné tržby za sledované období.....	32
Obrázek 3.2 Stanovení parametrů gama rozdělení modelu Pareto/NBD.....	35
Obrázek 3.3 Stanovení parametrů exponenciálního gama rozdělení modelu Pareto/NBD.....	36
Obrázek 3.4 Rozdělení pravděpodobnosti aktivity zákazníků.....	37
Obrázek 3.5 Frekvence opakovaných objednávek v trénovacím období.....	39
Obrázek 3.6 Frekvence opakovaných objednávek ve validačním období.....	40
Obrázek 3.7 Vztah mezi trénovacím a validačním obdobím.....	41
Obrázek 3.8 Srovnání skutečného a očekávaného počtu transakcí za celkové období	42
Obrázek 3.9 Srovnání skutečného a očekávaného kumulativního počtu transakcí ...	43
Obrázek 3.10 Návratnost investic dle různých scénářů.....	47
Obrázek 3.11 Rozdíly mezi ROI_1 a ROI_2.....	48
Tabulka 2.1 Přehled šetření v oblasti modelování CLV.....	16
Tabulka 3.1 Fragment průběžné přípravy datového souboru.....	29
Tabulka 3.2 Počet opakujících se objednávek.....	30
Tabulka 3.3 Popisná statistika základních kvantitativních proměnných.....	31
Tabulka 3.4 Odhad parametrů modelu Pareto/NBD.....	34
Tabulka 3.5 Charakteristiky konkrétního klienta s ID 1481.....	36
Tabulka 3.6 Paradox rostoucí četnosti.....	38
Tabulka 3.7 Data pro sestavení histogramu četností pro trénovací období.....	39
Tabulka 3.8 Data pro sestavení histogramu četností pro validační období.....	40
Tabulka 3.9 Srovnání odchylek skutečných a očekávaných dat.....	41
Tabulka 3.10 Zhodnocení trénovacího období.....	44
Tabulka 3.11 Zhodnocení validačního období.....	45
Tabulka 3.12 Fragment průběžné přípravy datového souboru	46

PŘÍLOHA 1

SKUPINA A									Udržení	Odchod	
Scenář	Sleva	Náklady na udržení	Výnosy z udržení	Ušlý zisk	Praděpodobnost aktivity	Pravděpodobnost odchodu	Očekavaný počet objednávek	CLV	ROI_1	ROI_2	Rozdíl mezi ROI
Optimistický	10%	258	2 453	129	0,95	0,05	3,7	2 583	850%	100%	750%
Realistický	15%	262	1 396	349	0,8	0,2	2,5	1 745	433%	25%	408%
Pesimistický	30%	398	928	398	0,7	0,3	1,9	1 326	133%	0%	133%

PŘÍLOHA 2

SKUPINA B									Udržení	Odchod	
Scenář	Sleva	Náklady na udržení	Výnosy z udržení	Ušlý zisk	Praděpodobnost aktivity	Pravděpodobnost odchodu	Očekavaný počet objednávek	CLV	ROI_1	ROI_2	Rozdíl mezi ROI
Optimistický	10%	86	601	258	0,7	0,3	2,7	859	600%	-67%	667%
Realistický	15%	105	350	350	0,5	0,5	2,2	700	233%	-70%	303%
Pesimistický	30%	134	134	312	0,3	0,7	1,4	445	0%	-57%	57%

PŘÍLOHA 3

SKUPINA C									Udržení	Odchod	
Scenář	Sleva	Náklady na udržení	Výnosy z udržení	Ušlý zisk	Praděpodobnost aktivity	Pravděpodobnost odchodu	Očekavaný počet objednávek	CLV	ROI_1	ROI_2	Rozdíl mezi ROI
Optimistický	10%	36	219	146	0,6	0,4	2,4	365	500%	-75%	575%
Realistický	15%	43	116	173	0,4	0,6	1,9	289	167%	-75%	242%
Pesimistický	30%	36	24	97	0,2	0,8	0,8	122	-33%	-63%	29%