

Vysoká škola ekonomická v Praze

Fakulta informatiky a statistiky

Katedra informačních technologií

Studijní program: Aplikovaná informatika

Obor: Informatika

Způsoby predikce dlouhodobé hodnoty zákazníka

BAKALÁŘSKÁ PRÁCE

Student : Martin Pešek

Vedoucí : Ing. Pavel Jašek

Oponent : doc. Ing. Ota Novotný, Ph.D.

2015

Prohlášení

Prohlašuji, že jsem bakalářskou práci zpracoval samostatně a že jsem uvedl všechny použité prameny a literaturu, ze které jsem čerpal.

V Praze dne 6. května 2015

.....

Martin Pešek

Poděkování

Děkuji vedoucímu práce panu Ing. Pavlu Jaškovi za jeho podporu a cenné rady a připomínky při zpracování této práce.

Abstrakt

Bakalářská práce se zabývá způsoby predikce dlouhodobé hodnoty zákazníka. Práce je zaměřena na predikce v non-contractual prostředí s využitím Pareto/NBD a BG/NBD modelů.

První část se věnuje konceptu dlouhodobé hodnoty zákazníka, možnostem jejího využití v Řízení vztahu se zákazníky, přínosům vyplývajících z tohoto konceptu a faktorům, které ovlivňují její predikce. Práce se dále zabývá životním cyklem zákazníka, vlivům datové kvality na realizaci predikcí a hodnotou zákazníka.

V další části jsou popsány základní předpoklady využití Pareto/NBD a BG/NBD modelů. Je zde popsáno, jaká data jsou nezbytná pro predikce a jaké výstupy tyto modely poskytují. Dále je představen proces predikce CLV a gamma-gamma model pro stanovení průměrné hodnoty transakcí.

V následující části práce je aplikována metodika CRISP-DM pro práci s datasetem internetového obchodu. Práce se tak zabývá jednotlivými fázemi metodiky. Podstatnou část tvoří zejména porozumění a příprava dat pro možnost jejich využití k predikcím dlouhodobé hodnoty zákazníka. Dále je proveden odhad parametrů modelů a následně srovnání výsledků predikcí zákaznického chování.

Závěrečná část práce se věnuje predikcím dlouhodobé hodnoty zákazníků a hodnoty zákaznické báze. Jsou zde porovnány simulované predikce pro vybrané zákazníky s rozdílným transakčním chováním a vytvořeny segmenty zákazníků podle jejich dlouhodobé hodnoty.

Práce tak poskytuje celý proces predikce dlouhodobé hodnoty zákazníků a srovnání Pareto/NBD a BG/NBD modelů s využitím reálných dat.

Klíčová slova

Dlouhodobá hodnota zákazníka, hodnota zákaznické báze, non-contractual prostředí, Customer Relationship Management, Pareto/NBD, BG/NBD

Abstract

This bachelor thesis covers methods of predicting customer lifetime value. The thesis is focused on predictions for non-contractual settings with use of Pareto/NBD and BG/NBD models.

The first part presents concept of customer lifetime value, possibilities of use in Customer Relationship Management, benefits from application of this concept and factors which have influence on predictions of customer lifetime value. This part of the thesis further covers customer lifecycle, influences of data quality on these predictions and customer value.

In the next part there are explained basic assumptions of Pareto/NBD and BG/NBD models. This part further explains about data requirements and which outputs are provided from these models. Following with the explanation of customer lifetime value prediction process and use of gamma-gamma spend model for calculating the average transactions values.

In the following part of the thesis CRISP-DM methodology is applied for use of the dataset from electronic retail. The thesis covers phases of this methodology. Main part is dedicated to understanding the data and their preparation for the following predictions of customer lifetime value. Then estimation of parameters is done and subsequently comparison of predicted customer behaviour.

The final part covers predictions of customer lifetime value and customer equity. This part is focused on simulation of predictions for selected customers with different transactional behaviour and creating segments of customers depending on their lifetime value.

The thesis thus provides the whole process of customer lifetime value prediction and comparison of Pareto/NBD and BG/NBD models with usage of real customer data.

Keywords

Customer lifetime value, customer equity, non-contractual settings, Customer Relationship Management, Pareto/NBD, BG/NBD

Obsah

1.	Úvod	1
1.1.	Téma práce	1
1.2.	Cíle práce.....	1
1.3.	Metody dosažení cílů práce	2
1.4.	Předpoklady a omezení práce.....	3
1.5.	Struktura práce	3
1.6.	Přínosy a výstupy práce	3
2.	Rešerše zdrojů	5
3.	Koncept customer lifetime value	8
3.1.	Customer lifetime value a customer equity	8
3.2.	Contractual a noncontractual prostředí	10
3.3.	Customer relationship management a CLV	12
3.3.1.	Zdroje dat pro predikci CLV	13
3.3.2.	Datová kvalita	14
3.4.	Životní cyklus zákazníka	16
3.5.	Hodnota zákazníka	17
3.6.	Přínosy a výhody predikce CLV pro podnik.....	19
4.	Modely	21
4.1.	Pareto/NBD model.....	21
4.1.1.	Pareto/NBD základní předpoklady a potřebná data	21
4.1.2.	Parametry a získané informace z Pareto/NBD modelu.....	21
4.2.	BG/NBD model.....	22
4.2.1.	BG/NBD základní předpoklady a potřebná data	22
4.2.2.	Parametry a získané informace z BG/NBD modelu.....	23
4.3.	Využití paretovských modelů k predikci CLV	23
4.3.1.	Gamma-gamma model základní předpoklady	23
4.3.2.	Výpočet CLV.....	24
5.	Praktická část – aplikování metodiky CRISP-DM.....	25
5.1.	Fáze porozumění businessu	25
5.2.	Porozumění datům	26
5.3.	Příprava dat.....	27

5.4.	Kalibrace modelů a odhad parametrů	29
5.5.	Zhodnocení modelů	30
5.5.1.	Pravděpodobnost aktivních zákazníků na konci kalibračního období	30
5.5.2.	Srovnání transakcí pro kalibrační a testovací období	31
5.5.3.	Gamma-gamma model.....	35
6.	Predikce CLV	37
6.1.	Individuální predikce.....	37
6.1.1.	Predikce CLV pomocí Pareto/NBD modelu	37
6.1.2.	Predikce CLV pomocí BG/NBD modelu	38
6.2.	Predikce CE.....	39
6.3.	Srovnání predikcí modelů pro jednotlivé zákazníky	39
6.4.	Zhodnocení možných dopadů predikcí na manažerská rozhodování	41
6.5.	Shrnutí.....	41
7.	Závěr	43
7.1.	Zhodnocení dosažení stanovených cílů	43
7.2.	Přínosy práce.....	44
7.3.	Možnosti využití a dalšího rozšíření práce.....	44
	Terminologický slovník.....	45
	Seznam literatury	46
	Seznam obrázků a tabulek	49

1. Úvod

1.1. Téma práce

Tato práce se věnuje způsobům predikce Customer lifetime value (dlouhodobé - „celoživotní“ hodnoty zákazníků), výhodám a možným dopadům plynoucím z těchto predikcí na manažerská rozhodování různých úrovní. Práce se zabývá predikcí dlouhodobé hodnoty zákazníků v non-contractual prostředí (bez smluvního vztahu), kde vztah mezi podnikem a zákazníkem není řízen pomocí smluvních vztahů a nelze tak přesně stanovit, kdy zákazník tento vztah ukončí.

Konceptem Customer lifetime value (dále jen CLV) se zabývá řada publikací a akademických článků (viz kapitola 2). Práce se věnuje možnostem využití tohoto konceptu v kontextu CRM (Customer Relationship Management), s jakými problémy se predikce CLV potýká a jaké výhody přináší využití jeho využití v prostředí CRM.

Dále je v této práci vysvětlen proces predikce CLV: jaké faktory ovlivňují samotný způsob výpočtu CLV, tedy jaký model pro predikci využít a jaké faktory ovlivňují přesnost daného modelu pro výpočet CLV.

1.2. Cíle práce

Hlavní cíle práce jsou:

1. analyzovat využívané modely pro predikci CLV,
2. prakticky otestovat predikce CLV na zákaznických datech,
3. kriticky zhodnotit tyto predikce k možným dopadům na manažerská rozhodování v různých úrovních řízení.

Dílčí cíle jsou:

- provést rešerši zdrojů pro získání přehledu o konceptu CLV a nejčastějších způsobech predikce za daných situací
- analyzovat vhodnost vybraných modelů
- aplikace zavedené metodiky na práci s daty pro predikce CLV v praktické části práce

První cíl se bude zabývat problémem, jaký model použít pro predikci CLV, jaké jsou jeho výhody a nevýhody a jaká omezení jsou spjata s využitím tohoto modelu. Následně bude posouzena kvalita tohoto modelu. K dosažení tohoto cíle bude nezbytné omezit tuto práci na predikování CLV za určité situace, aby nedošlo k překročení jejího rozsahu (viz kapitola 1.4.).

Dříve než bude možné splnit tento cíl, bude potřeba provést rešerši zdrojů věnujícím se dané tématice. Splněním dílčího cíle bude získán přehled o tématice a budou identifikovány nejčastější způsoby predikce CLV, tedy jaké modely se v daných situacích používají.

Splněním tohoto cíle bude možné analyzovat identifikované modely a posoudit vhodnost jejich využití, s jakým omezeními jsou spojeny.

Druhý cíl povede k aplikování modelů identifikovaných v předchozím cíli na zákaznických datech. Tomuto cíli bude věnována praktická část práce. Na základě využití dat budou provedeny predikce CLV. Získané hodnoty budou porovnány a bude tak možno porovnat použité modely a posoudit jejich výsledky.

Pro naplnění druhého cíle bude také nezbytné aplikovat metodiku CRISP-DM¹ (viz kapitola 5). Bude nutno se věnovat jednotlivým fázím práce s daty podle této metodiky, které budou dále popsány. Jedná se o následující fáze: porozumění business problému, porozumění datům, příprava dat, modelování, vyhodnocení výsledků a jejich využití.

Následně bude plněn třetí cíl, kdy na základě provedených predikcí bude možné zhodnotit důsledky na manažerská rozhodování. Zde bude možno posoudit dlouhodobou hodnotu zákazníků a kriticky zhodnotit jak tyto hodnoty ovlivní manažerská rozhodování různých úrovní managementu. Tedy zda by měl podnik usilovat o udržení zákazníků, na jaké zákazníky se zaměřit a podobně. Dále pomocí CLV lze stanovit CE (customer equity), která slouží jako metrika pro kvantifikování celé zákaznické báze podniku (viz kapitola 3.1).

1.3. Metody dosažení cílů práce

Provedením rešerše v další kapitole bude z příslušných zdrojů nejprve získán teoretický přehled o konceptu CLV a její predikci. Na základě získaných informací bude vypracována kapitola 3, která se bude zabývat základními pojmy pro tuto tematiku. Dále bude řešen vztah CLV k CRM, faktory ovlivňující predikce CLV a možné přínosy při využití CLV pro podnik.

Následně bude identifikován způsob predikce CLV pro non-contractual prostředí pomocí vhodných modelů. Na základě těchto modelů budou provedeny predikce CLV s využitím zákaznických dat.

K využití dat bude nutné použít metodiku CRISP-DM, kterou popisuje například Chapman et al. (2000). Před použitím dat se bude nezbytné věnovat fázi porozumění dat, přípravě dat a zbylým částem této metodiky

Nakonec ze všech předchozích kapitol práce budou získány informace, pomocí kterých budou vyhodnoceny dopady na manažerská rozhodování a jak ovlivní predikce CLV vztah podniku k jeho zákazníkům.

¹ Tato zavedená metodika pro práci s daty byla vybrána na základě zkušeností autora s touto metodikou.

1.4. Předpoklady a omezení práce

Tato práce je omezena na predikce CLV v situaci non-contractual prostředí. Důvodem omezení je dodržení rozsahu bakalářské práce a zejména dostupnost a charakter zákaznických dat využívaných v praktické části práce.

Toto prostředí je charakteristické tím, že vztah zákazníků s podnikem není řízen žádným smluvním vztahem. Zákazníci uskutečňují nákupy náhodně v čase, často i nakupují u ostatních podniků a opětovně se navrací k původnímu podniku. Proto podniky nepovažují nákup zákazníka u jiného podniku za ztrátu tohoto zákazníka. Jelikož zákazník může nakupovat od daného podniku s různou frekvencí, je potřeba v predikci CLV vyřešit otázku, zda je zákazník stále aktivní a jakým zákazníkům věnovat pozornost. Tato práce bude omezena na modely Pareto/Negative Binomial Distribution a Beta Geometric/Negative Binomial Distribution (dále jen Pareto/NBD a BG/NBD), které jsou vhodné pro použití v tomto kontextu.

1.5. Struktura práce

Práce je rozdělena na teoretickou a praktickou část. V jednotlivých kapitolách je na začátku stanoveno čemu se daná kapitola věnuje, jaký je její cíl a v závěru jsou stanoveny hlavní poznatky z dané kapitoly.

V teoretické části je provedena rešerše zdrojů (viz kapitola 2.). Dále je zde vysvětlen koncept CLV (viz kapitola 3.) a jeho vztah k CRM. Nakonec jsou v teoretické části představeny vybrané modely pro predikci CLV (viz kapitola 4.).

Praktická část se věnuje zhodnocení modelů na datasetu internetového obchodu a simulaci predikcí CLV pomocí modelů vybraných v teoretické části. K tomu je nezbytné aplikování metodiky na využitá data, kterou je metodika CRISP-DM. Aplikování této metodiky je věnována kapitola 5. Dále jsou již provedeny predikce CLV a jejich zhodnocení.

Závěr práce se věnuje kritickému zhodnocení získaných hodnot CLV a jejich dopadům na různé úrovně manažerských rozhodování.

1.6. Přínosy a výstupy práce

Čtenářům, kteří se tématikou predikce CLV dosud nezabývali, poskytne tato práce především teoretický pohled na koncept CLV. Jistě tak ocení znalosti získané z kapitoly 3. Dozví se o tom, jaké výhody přináší predikce CLV a jaké modely lze využít v non-contractual prostředí.

Pro akademické čtenáře bude práce přínosná zejména z pohledu, jak zapadá koncept CLV do kontextu CRM. Dále práce poskytne příklad aplikování metodiky CRISP-DM na praktickém příkladu.

Praktikové, kteří se už s tematikou zabývají, naleznou přínosy této práce především v částech věnovaných analýze modelů, kde budou vybrané modely zhodnoceny. Výsledky této práce budou moci případně porovnat s jejich vlastními výsledky. Přínosem budou faktory ovlivňující kvalitu modelů, posouzení výsledků predikcí pomocí těchto modelů.

Výstupem řešerše zdrojů bude teoretický souhrn, kde budou identifikovány modely vhodné pro non-contractual prostředí a také získány potřebné informace pro tuto práci. Dalším výstupem práce bude teoretický základ věnující se konceptu CLV a příslušným pojmům v kapitole 3. Dalšími výstupy práce bude posouzení identifikovaných modelů, simulace predikcí CLV a nakonec souhrn vlivů CLV na manažerská rozhodování jednotlivých úrovní.

2. Rešerše zdrojů

Cílem této kapitoly je provést rešerši, kde z jednotlivých zdrojů budou identifikovány přínosné informace pro téma této bakalářské práce. U jednotlivých zdrojů je vysvětleno, z jakého pohledu autoři řeší danou problematiku a jaké přínosy nabízejí.

Machander, Aleš. *Aplikace CVM v telekomunikační společnosti*. Praha, 2012. Diplomová práce. Vysoká škola ekonomická.

Tato práce se zabývá konceptem CVM (customer value management) a retenčním procesem v prostředí telekomunikační společnosti. Jedná se tedy o contractual prostředí, pro které je typický retenční proces, zatímco tato práce se bude věnovat non-contractual prostředí, v kterém se podle Fader (2012) nemá smysl zabývat tímto procesem. Dále se autor v práci zabývá obecným model CLV a životním cyklem zákazníka. Životní cyklus zákazníka je popisován v kontextu telekomunikační společnosti, v této práci mu bude věnována kapitola, která bude řešit obecně jeho vlivy na výpočet CLV.

Pavel, Jan. *Customer Intelligence v prostředí elektronického obchodu*. Praha, 2012. Diplomová práce. Vysoká škola ekonomická.

Autor této práce se zabývá konceptem CI (customer intelligence) v prostředí internetového obchodu. Zmiňuje se také o kvalitě dat v CRM systému ve zkoumaném podniku a jejím vztahu ke CI a webové analytice. V této bakalářské práci bude řešena kvalita dat zejména ve vztahu k predikování CLV. Budou tak popsány základní požadavky na datovou kvalitu a jejich vliv na CLV. Autor se také částečně zabývá retencí zákazníků v oblasti internetového obchodu, která má vliv na zvyšování CLV zákazníků.

Platzer, Michael. *Stochastic Models of Noncontractual Consumer Relationships*. Vídeň, 2008. Diplomová práce. Vienna University of Economics and Business Administration.

Autor této práce analyzuje modely predikce zákaznického chování v non-contractual prostředí. Cílem práce bylo analyzovat a vybrat vhodné modely pro toto prostředí v rámci soutěže pořádané Direct Marketing Educational Foundation v roce 2008, odkud autor získal dataset, na kterém provádí analýzu modelů. Autor srovnává jednak NBD model a Pareto/NBD a BG/NBD modely, které z něho vycházejí. Ty modely budou také analyzovány v této bakalářské práci a budou s nimi navíc provedeny predikce CLV a CE. Kromě těchto modelů autor srovnává CBG/NBD a jednotlivé varianty CBG/CNBD-k modelu. Autor aplikuje tyto modely na dataset a srovnává zákaznické chování predikované těmito modely, jako je například frekvence zákaznických transakcí a očekávaný počet transakcí realizovaných zákazníky. Nejpřínosnější částí práce je jednak provedení predikcí zákaznického chování, ale také získané výsledky. Nejlepší výsledků podle autora dosáhly modely typu CBG/CNBD-k. Jednou ze zajímavostí je také porovnání výsledků komplexních modelů se základním lineárně regresním modelem. Ten nezřídka poskytuje poměrně kvalitní predikce (i v případě datasetu zkoumaného autorem). Nicméně jak autor dodává, tento model sice poskytuje poměrně přesné predikce, ale nedosahuje kvalitu komplexních modelů.

Gupta, S. et al Modeling Customer Lifetime Value. 2006

Autoři článku poskytují ucelený přehled modelů pro predikce CLV. Kategorizují modely pro určité situace, podle vztahu s podnikem. Nabízejí pohled na vztah marketingových programů k získávání, udržení a rozšiřování zákazníků, které následně ovlivňuje CLV a CE hodnoty tvořící celkovou hodnotu podniku. V rámci tohoto vztahu autoři doporučují využití jednotlivých druhů modelů k predikci CLV. Autoři tak kategorizují modely zkoumající vliv marketingových plánů a kampaní na získání, udržení a rozšíření zákazníků. Další modely zkoumají vztah jednotlivých komponent CLV (např. vztah mezi akvizicí a retencí zákazníků), řeší vztah CLV či CE k hodnotě podniku. Mezi často používané modely patří RFM modely, které používá také například Pavel (2012), ke skórování zákazníků. Pro účely této bakalářské práce jsou však nejpodstatnější pravděpodobnostní modely. Tyto modely jsou vhodné zejména v non-contractual prostředí, kde se přesně nedá stanovit, zda je zákazník stále aktivní v určitém období. Pravděpodobnostní modely tak pro zákazníky predikují pravděpodobnost aktivity a očekávaný počet transakcí. Do této skupiny patří Pareto/NBD model. Jak bylo zmíněno v kapitole 1.4, tato práce se bude věnovat dále podrobněji tomuto typu modelů.

Berger, Paul D. a Nada I. Nasr. Customer lifetime value: Marketing models and applications. 1998

Autoři tohoto článku se věnují řadě modelů, nejprve se věnují obecnému způsobu vyjádření CLV a CE. Zde zajímavě popisují jaké komponenty je nutno zahrnout do výpočtu CE. Na rozdíl od definice podle Fader (2012), je nutno do této hodnoty zahrnout i jednotlivé náklady spojené s udržení zákazníků, náklady spojené s transakcemi apod. Pro účely této bakalářské práce bude však využita zjednodušená definice (viz kapitola 3.1). Dále se autoři věnují zákaznickému chování, na jehož základě vymezují retenční modely a migrační modely. V tomto článku je zkoumáno, jak často probíhají prodeje produktů. Modely pracují s hrubou marží zákazníků, od které jsou odečítány náklady spojené se zákazníkem. Přestože modely jsou řešeny nezávisle na kontextu, tento článek nabízí řadu číselných příkladů, které byly přínosné pro porozumění tématice pro tuto práci.

Fader, Peter S. a Bruce G.S. Hardie. Probability Models for Customer-Base Analysis. 2009

Autoři článku se věnují pravděpodobnostním modelům pro predikci CLV, dále také řeší užití těchto modelů ve vztahu k contractual a non-contractual prostředí. Přínosem pro tuto práci je tedy druhá z těchto variant. Jak již bylo zmíněno výše v článku (Gupta, S. et al, 2006), vhodným pravděpodobnostním modelem pro non-contractual prostředí je Pareto/NBD model, který je i v tomto článku rozebírán. Dále je zde představen BG/NBD model, který navazuje na předchozí model. Pro praktickou část práce tak budou vybrány tyto dva paretovské modely. Bude tak možné provést srovnání výsledků predikcí těchto modelů.

Fader, Peter. *Customer centricity: Focus on the Right Customers for Strategic Advantage*. Philadelphia, 2012.

V tomto zdroji se autor věnuje vztahu podniku ke svým zákazníkům, především je zde rozebírán vztah customer centricity a jeho přínosy pro podnik. Autor se věnuje oproti Gupta et al. (2006) a Berger (1998) pojmům CLV a CE jen z teoretického hlediska. Dále je rozebrán vztah CRM k CLV a ke customer centricity. Informace z této knihy poslouží zejména teoretické části této práce věnující se konceptu CLV v kapitole 3. Kromě těchto pojmů je také v knize řešen problém na jaké zákazníky by se podnik měl zaměřit, což je jednou z otázek, které podnik čelí po tom, co je predikována CLV zákazníků.

Novo, Jim. *Drilling Down: Turning Customer Data Into Profits with a Spreadsheet*. Saint Petersburg, 2004.

Zajímavou částí tohoto zdroje jsou zejména části, kde se autor věnuje dlouhodobé hodnotě zákazníka. Je zde řešen vztah životního cyklu a hodnoty zákazníka k CLV. Autor navíc řeší vztah CRM a CLV, jak tato hodnota může přispět k návratnosti investice z tohoto systému. Tento zdroj je přínosný zejména pro získání celkového přehledu o konceptu CLV.

MCCARTHY, Daniel a Edward WADSWORTH. *Buy 'Til You Die - A Walkthrough*, 2014.

Tento zdroj bude využit hlavně v praktické části bakalářské práce. Autoři této příručky k balíčku BTYD popisují proces přípravy dat a aplikování modelů Pareto/NBD, BG/NBD a BG/BB pro predikce zákaznického chování. Zajímavou částí tohoto zdroje jsou právě kapitoly věnující jednotlivým fázím prvních dvou modelů. Autoři vysvětlují využití příslušných funkcí z balíčku BTYD pro implementaci těchto modelů s využitím programovacího jazyka R.

BAGGOTT, Matthew. *Predicting Customer Behavior with R: Part 1*, 2013.

Autor tohoto zdroje popisuje implementaci Pareto/NBD modelu pomocí programovacího jazyka R. Na rozdíl od Platzner (2008) využívá k predikcím veřejný dataset CDNOW, který je využíván i v dalších článcích věnujících se dané problematice. Autor popisuje implementaci tohoto modelu pro predikci zákaznického chování. Vysvětluje proces přípravy dat, odhad parametrů, využití modelu pro individuální predikce očekávaného počtu transakcí a pravděpodobnosti, zda je zákazník aktivní v určitém období. Kromě samotných predikcí je provedeno i zhodnocení modelu v porovnání s reálnými daty s CDNOW. Autor poskytuje i příslušné skripty obsahující kód využitelný v programovacím jazyku R. Tyto skripty tak realizují výpočty tohoto modelu nad zmíněným datasetem. Bude je tak možné upravit pro využití v této bakalářské práci pro analyzovaný dataset internetového obchodu.

3. Koncept customer lifetime value

Cílem této kapitoly je vymezit základní pojmy spojené s tématem CLV a její predikce. Kapitola se věnuje konceptu CLV a výhodám, které přináší její predikce. Dále se kapitola zabývá faktory, které mají vliv na kvalitu a přesnost predikcí. Je v ní popsán vztah CRM a CLV, jak tento systém poskytuje data pro predikci CLV a jakým způsobem lze využít konceptu CLV v tomto prostředí.

Význam CLV spočívá především v možnosti vyjádření hodnoty zákazníků do budoucna. Pomocí této hodnoty je možné určit profitabilitu zákazníků. Slouží také jako metrika pro řadu manažerských rozhodování. Jejím využitím lze lépe rozhodnout do kterých zákazníků investovat za účelem maximalizace ROI (return of investment - návratnost investice).

3.1. Customer lifetime value a customer equity

Přestože definice CLV jsou si ve značné míře podobné, lze mezi nimi v různých zdrojích nalézt drobné odlišnosti. Tyto odlišnosti právě odrážejí i myšlenky autorů, jaké komponenty nebo náklady by měly být započítávány do CLV.

Definice podle Fader (2012) je následující: „*Customer lifetime value je současná hodnota budoucích net cash flows spojených s konkrétním zákazníkem.*“

Dále například Gupta et al.(2006) používá definici: „*Customer lifetime value je současnou hodnotou všech budoucích zisků získaných od zákazníka za jeho život ve vztahu s podnikem.*“

Přestože tedy samotná definice CLV není na první pohled až tolik složitá, existují rozdílné názory, jak nahlížet na CLV a jak ji predikovat. Jedním z rozporů je, zda zahrnout do CLV náklady spojené s akvizicí zákazníka. V článku Berger (1998) autoři nezahrnují tyto náklady do definice a výpočtu CLV, zatímco Jain (2002) tyto náklady bere za nezbytnou součást CLV.

Fader (2012) jeho definici dále rozvíjí o čtyři faktory, které je nutné vzít v potaz pro predikci CLV.

1. Jde o dopředu nahlízející koncepci
 - CLV tedy nezahrnuje profitabilitu zákazníka v minulosti
2. využití jen relevantních dat
 - pro predikci CLV by měly být použity jen ty náklady, které jsou úzce spjaté se zákazníkem
3. výpočet CLV je pouze odhadem
 - vypočtená hodnota CLV udává predikovanou hodnotu, přestože využitím kvalitního modelu dostaneme konkrétní hodnotu CLV, tak tato hodnota se může lišit od reálné hodnoty
4. odlišnosti podle druhu podnikání

- predikce CLV se liší pro non-contractual a contractual prostředí (v contractual prostředí je možné mluvit o retenčním procesu a ukončení vztahu mezi zákazníkem a podnikem, zatímco v non-contractual prostředí tyto fáze neprobíhají)

Důležité faktory pro porozumění konceptu CLV jsou uvedeny v prvním a čtvrtém bodu. Druh podnikání je základním faktorem ovlivňující přístupy predikce CLV a je mu věnována podrobněji následující kapitola 3.2. Jak bylo uvedeno v úvodu kapitoly, CLV je využívána mimo jiné jako metrika zákaznické hodnoty, jejíž charakteristikou je právě to, že se jedná metriku nahlížející i do budoucna. Kumar, V. (2008) popisuje rozdíl mezi CLV a dalšími běžně využívanými metrikami, jako jsou RFM, PCV (past customer value), SOW (share-off-wallet), které oproti CLV tuto vlastnost nemají. Tyto metriky vycházejí z již pozorovaného zákaznického chování, ale nezabývají se tím, zda zákazník bude realizovat transakce v budoucnu. RFM koncept vychází z toho, kdy zákazník byl naposledy aktivní, s jakou frekvencí realizoval transakce a jakou peněžní částku utratil od minulosti až do současnosti. PCV vyjadřuje celkovou hodnotu zákazníka od minulosti do současnosti. SOW se zabývá tím, jaké měl zákazník preference mezi jednotlivými značkami v minulosti. CLV tedy naopak zahrnuje pravděpodobnost, s jakou může podnik očekávat, že bude zákazník aktivním v budoucnu, jakou hodnotu zákazník přinese podniku v budoucnu a jaké náklady jsou spojeny s daným zákazníkem. Na základě těchto informací může podnik rozhodovat o tom, kteří zákazníci jsou pro podnik profitabilnější a přizpůsobit tak marketingové plány a kampaně.

Zbývající dva faktory jsou neméně důležité pro predikci CLV a manažerská rozhodování, neboť ovlivňují přesnost predikcí a to, jak podnik na tento koncept nahlíží.

Customer equity

Dalším důležitým pojmem spojeným s konceptem CLV je customer equity, která slouží k ohodnocení zákaznické báze podniku. Lze ji mimo jiné použít jako ukazatel ekonomického růstu podniku. Existují různé definice jak tuto hodnotu vyjádřit.

Fader (2012) definuje CE následovně: „*Customer equity je suma dlouhodobých hodnot zákazníků napříč celou zákaznickou bází podniku.*“

Vyjde-li se z této definice, lze jednoduše vyjádřit CE součtem jednotlivých hodnot CLV. Definice nabízí snadnější postup jak vypočítat CE, a proto bude využita i v teoretické části této práce.

O něco složitější forma pro vyjádření CE, je například definována v článku Berger (1998). Podle tohoto zdroje je potřeba do výpočtu CE zahrnout navíc i náklady spojené se získáním a udržením zákazníka, marži za jednotlivé transakce a další proměnné. Dále je nezbytné do CE zahrnout i CLV hodnoty nově získaných zákazníků.

Customer ekvity je tedy důležitým ukazatelem, odrážející souhrnný stav jednotlivých CLV hodnot. Podnik by měl směřovat k maximalizaci této hodnoty. CE je důležitou metrikou, která odráží stav celého podniku.

CLV naopak odráží hodnotu jednotlivých zákazníků, pomocí které lze určit profitabilitu těchto zákazníků a lze ji využít k segmentaci zákazníků. Pomocí CLV podnik lze zákazníky rozdělit do skupin podle jejich CLV. Lze tak segmentovat zákazníky například s vysokou, střední hodnotou až po zákazníky s téměř nulovou hodnotou CLV. Podnik pak může lépe cílit marketingové kampaně a strategie na příslušný segment zákazníků.

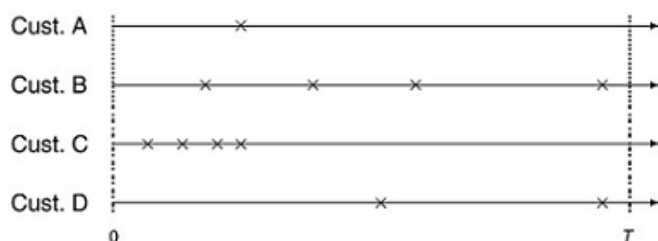
3.2. Contractual a noncontractual prostředí

V konceptu CLV a její predikce jsou charakteristické dva směry, které se odvíjí od druhu podniku a jeho vztahu k zákazníkům. Pro predikci CLV je důležité rozlišit, o kterou situaci se jedná. Jednou z nich je contractual prostředí, kdy je vztah řízen pomocí nějaké formy smluvního vztahu. Zákazníci často nakupují a využívají produktů a služeb podniku ve smluvené době. Hlavním kritériem této situace je však to, že podnik ví, kdy ztrácí svého zákazníka.

V non-contractual prostředí tato informace není podniku známa, a proto je podnik naopak nucen odhadovat, je-li zákazník stále aktivní. Fader (2012) zmiňuje, že v tomto prostředí by bylo zcestné zabývat se retencí zákazníků a zkoumáním, zda se zákazník chystá přejít k jinému podniku. Proto je důležité si uvědomit to, že retence zákazníka v non-contractual prostředí vůbec neprobíhá, a proto nelze také řešit odchod zákazníka od podniku.

Jak již bylo zmíněno v kapitole 1.4, tato práce je zaměřena na non-contractual prostředí. Jedná se o situaci, kdy vztah podniku se zákazníky není stanoven pomocí žádné smluvního vztahu, zákazníci u podniku nakupují s náhodnou frekvencí. Rovněž v této situaci zákazníci často nakupují u různých podniků, přesto takové chování zákazníků není bráno jako ukončení jejich vztahu s podnikem. Existuje totiž šance, že zákazník bude u podniku nakupovat v budoucnu. Jde o tzv. always-a-share vztah zákazníků s podnikem. Přesto podnik řeší otázku, zda je konkrétní zákazník stále aktivní.

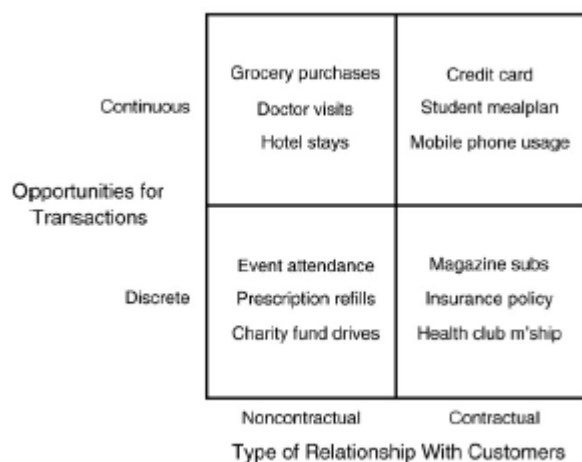
Fader (2009) popisuje různé varianty transakčního chování zákazníků. Obr. 1 znázorňuje tyto varianty provedených transakcí (nákupů), z něhož lze pozorovat různé frekvence nákupů během časového horizontu (0; T].



Obr. 1 - Transakční historie (Fader, 2009)

Zákazník A realizoval pouze jeden nákup a podnik může odvodit, že existuje poměrně malá šance, že by nakoupil v budoucnu (po bodu T). Zákazníci B a C nakupovali čtyřikrát, přesto u zákazníka C uplynula delší doba od jeho poslední transakce, proto lze u něho očekávat nižší pravděpodobnost budoucího nákupu. Z uplynulé doby od poslední transakce tak podnik může odvodit šanci budoucího nákupu a z množství realizovaných transakcí lze odvodit potencionální množství transakcí v budoucnu.

Fader (2009) uvádí další dvě rozdělení pro contractual i non-contractual prostředí. Jedná se o klasifikaci transakcí na diskrétní a spojitě rozdělení. Diskrétní rozdělení transakcí je takové, kdy transakce probíhají pouze ve stanovené době. Spojitě rozdělení naopak představuje situaci, kdy zákazník realizuje transakce neomezeně k určité době. K realizaci transakcí tak může dojít kdykoliv. Tato rozdělení zachycuje Obr. 2.



Obr. 2 - Klasifikace vztahů se zákazníky (Fader, 2009)

Před přistoupení k predikci CLV je nutné stanovit, do jakého ze čtyř kvadrantů podnik s jeho vztahem k zákazníkům spadá. Na základě této klasifikace se odvíjí způsob predikce CLV i výběr vhodného modelu. Z výše uvedeného transakčního chování zákazníků, lze na druhé straně odvodit, jakou CLV hodnotu přinese zákazník podniku a zda je stále aktivním zákazníkem.

3.3. Customer relationship management a CLV

Tato kapitola řeší vztah CRM a CLV a vysvětluje proč aplikovat tento koncept v CRM.

V současné době s rozvojem IS/ICT jsou kladeny stále větší požadavky na CRM a jeho využití. Podniky získávají ohromné množství zákaznických dat z těchto systémů, v závislosti na velikosti podniku a délce jeho podnikání dosahují úrovně big data. Tato data je nutné dále analyzovat, má-li z nich podnik získat znalosti o svých zákaznících a s tím i konkurenční výhodu.

Právě v kombinaci využití dat ze CRM a predikce CLV zákazníků může podnik získat konkurenční výhodu a podpořit vztah zákazníků a jejich loajalitu. Diener (2013) popisuje právě, jak CRM bez využití CLV a big data sice podniku výhody samotného CRM, ale podnik tak ztrácí možnost zvyšovat loajalitu zákazníků a udržení si konkurenční výhody. Využitím predikce CLV podnik může získat řadu dalších výhod (viz kapitola 3.4.).

Pořizování CRM

Podnik, který dosud nemá CRM, čelí otázkám, zda takový systém potřebuje a jaká bude návratnost investice z jeho pořízení. Vyřešením těchto otázek ovlivňuje úspěšnost či neúspěšnost implementace CRM. Před samotným zvažováním o pořízení CRM, by měl každý podnik získat určité znalosti o svých zákaznících a jejich chování. Novo (2004) uvádí následující otázky, které by měly být zodpovězeny.

1. Identifikování nejprofitabilnějších zákazníků.
2. Kolik zákazníků odešlo od podniku za poslední časový horizont.
3. Poměr mezi nejprofitabilnějšími zákazníky, kteří odešli od podniku a zákazníky, kteří zůstali.
4. Určení ušlého zisku z nejprofitabilnějších zákazníků, kteří odešli.
5. Výše nákladů pro udržení těchto zákazníků.
6. Určení návratnosti investice.

Pro identifikování nejprofitabilnějších zákazníků musí být podnik schopen stanovit výši nákladů spojených s transakcemi se zákazníky, marži, náklady spojené se servisními službami apod. Na základě těchto informací je podnik schopen identifikovat jeho nejlepší zákazníky. Aby byly tyto informace k dispozici, musí být evidována data z jednotlivých transakcí.

Dále je nutné stanovit, kolik z nejprofitabilnějších zákazníků od podniku odešlo, respektive přestalo být aktivní během určitého období. Určit, zda zákazník odešel od podniku, může být někdy obtížnější. Závisí právě na druhu podnikání (viz kapitola 3.2). Proto je v non-contractual prostředí také potřeba identifikovat zákazníky, kteří se mohou po delší době neaktivity stát opět aktivními a případně nejprofitabilnějšími.

Dalším krokem je porovnat tyto zákazníky se zákazníky, kteří od podniku neodešli, respektive se nestali neaktivními. Je tak třeba analyzovat zákaznická data a identifikovat mezi nimi rozdíly, které vedly k odchodu zákazníků.

Z předchozích informací lze již zjistit výši ušlého zisku za zákazníky, kteří od podniku odešli. Následně je třeba určit, kolik by stálo udržení těchto zákazníků.

Na základě těchto informací je možné lépe stanovit návratnost investice. K zodpovězení předchozích otázek je možné aplikovat koncept CLV, který pomáhá identifikovat profitabilní zákazníky, stanovit jejich hodnotu a predikovat ji i do budoucnosti.

3.3.1. Zdroje dat pro predikci CLV

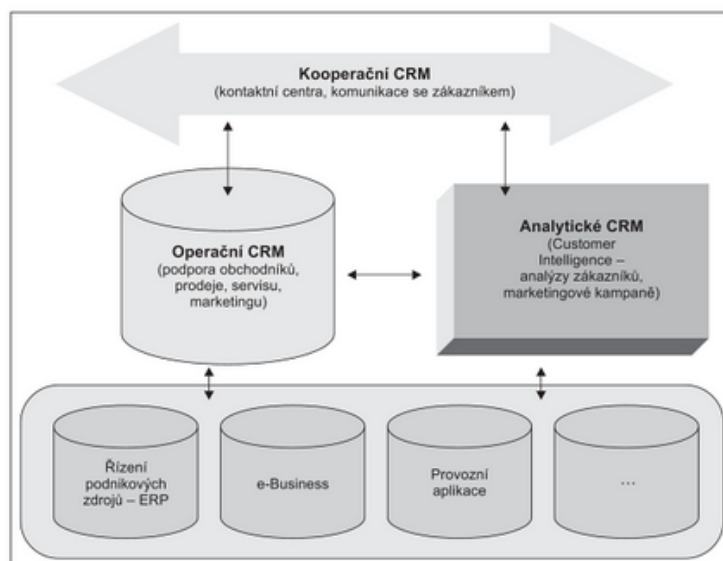
K predikci CLV zákazníka se využívají data, která jsou vytvářena transakcemi daného zákazníka. Tato data jsou vytvářena při obchodních procesech podniků a jsou ukládána v CRM systému.

Gála et. al (2009) dělí CRM na tři základní oblasti:

1. Operační CRM
2. Kooperační CRM
3. Analytické CRM

První dvě oblasti mají transakční charakter, zahrnují data získaná z obchodních případů, komunikace se zákazníkem apod. Tyto oblasti CRM tak přinášejí podniku potřebná data o zákaznících, z kterých lze vypočítat CLV. Následně získané hodnoty CLV jsou využívány v analytickém CRM. Lze tak využít predikce pro různá manažerská rozhodování jako například rozpoznat zákazníky s největší profitabilitou a zaměřením se na tyto zákazníky, určení jaké části rozpočtu investovat do jednotlivých segmentů zákazníků, porovnat profitabilitu různých marketingových kampaní a strategií apod.

Na Obr. 3 je shrnuta základní architektura CRM podle Gála et. al (2009), která představuje jednotlivé oblasti CRM a lze na ní lépe popsat vztah jednotlivých oblastí CRM a jejich souvislosti s konceptem CLV.



Obr. 3 - Architektura CRM - Gála et. al (2009)

Operační část CRM zprostředkovává obchodní případy, vytváření marketingových plánů a kampaní servisní služby pro zákazníky apod. Zejména pak marketingové plány a kampaně a servisní služby podporují CLV. Kvalita těchto aktivit vede k větší loajalitě zákazníků a jejich spokojenosti s nabídkou služeb a produktů podniku, čímž se zvětšuje i CLV zákazníků.

Kooperační část CRM zahrnuje zejména komunikaci se zákazníkem skrze různé komunikační kanály (pošta, mobilní komunikace, webové aplikace aj.) a je obvykle řízena call-centry. Právě v call-centrech by bylo možné aplikovat koncept CLV. Za předpokladu, že by podnik znal CLV hodnoty svých zákazníků, mohli by tak operátoři na call-centrech přistupovat individuálně k jednotlivým zákazníkům. Operátoři by věděli, kterým zákazníkům by se měli snažit nejvíce vyhovět a mohli by přizpůsobit délku hovorů pro zákazníky podle jejich CLV hodnoty. Pomocí optimálního alokování času na call-centrech by bylo možné zlepšit jejich efektivitu a zvýšit spokojenost zákazníků. Aplikování konceptu CLV tak může přispět ke zlepšení kooperační části CRM.

Analytická část oproti předchozím oblastem CRM slouží zejména k využití informací o zákaznících získaných z předchozích dvou oblastí. Právě v analytickém CRM lze mimo jiných aktivit využít koncept CLV pro určení hodnoty zákazníka a její predikce do budoucnosti. Lze predikovat zákaznické chování a využít CLV jako metriky pro měření hodnoty jednotlivých zákazníků a také k určení CE jako metriky po hodnotu podniku.

3.3.2. Datová kvalita

Datová kvalita je jednou z příčin, která může mít vliv na přesnost predikce CLV zákazníků, proto je potřeba neopomenout ji v této práci a hlavně by neměla být opomíjena i v kontextu predikce CLV. Právě zákaznická data jsou pro podniky velmi důležitým zdrojem informací a je potřeba zajistit kvalitu těchto dat.

Některé z požadavků na kvalitu dat definuje Olson (2003), který definuje kvalitní data, jako ta, která jsou:

1. přesná,
2. včasná,
3. relevantní,
4. úplná,
5. srozumitelná,
6. důvěryhodná.

Prvním požadavkem na kvalitní data je přesnost dat, kdy je potřeba zajistit formu ukládaných dat. Je důležité, aby data byla ukládána v jednotné podobě a zamezit případným nekonzistencím a duplicitám v datech. Tato vlastnost dat stanovuje, jak se data liší od skutečnosti, tedy jednotlivá chybná data snižují přesnost dat. Čím přesnější data budou využita pro predikci CLV, tím přesnější tedy bude i odhadovaná CLV. Důsledkem toho využití predikce CLV pro manažerská rozhodování přinese podniku o to větší výhody.

Dále je důležité, aby data byla včasná respektive aktuální. Tento problém je na jedné straně spjat s různými kontaktními údaji zákazníků, které se mohou měnit. Na druhé straně to mohou být data, ze kterých se predikuje CLV. Je nezbytné, aby podnik získal včas i ta nejnovější data. Nekvalitou z toho hlediska se rozumí, když podnik ještě nemá k dispozici nově vzniklá data z transakcí zákazníka a nemůže jich využít pro predikci CLV. Proto je potřeba zajistit pravidelné aktualizace dat.

Relevance dat je dalším z požadavků na kvalitu dat. Podnik o svých zákaznících uchovává mnoho rozdílných dat, proto musí vědět, která data přísluší ke kterému zákazníkovi. Je tedy nezbytné, aby pro predikci CLV určitého zákazníka byla využita pouze data spojená s tímto zákazníkem.

Čtvrtým z požadavků na datovou kvalitu je, aby data byla úplná. Podnik by měl mít o jednotlivých zákaznících všechna data. Právě i v CRM systémech lze pozorovat situace, kdy podnik může mít všechna data pro určité zákazníky a zároveň pro ostatní některá z dat nemusí být k dispozici. Tento problém je nutné vyřešit při predikci CLV zákazníků, u kterých některá data chybí. Je proto zapotřebí následně většího úsilí při přípravě dat před predikcí CLV a stanovit postup, jak zacházet se zákazníky s neúplnými daty.

Dalším požadavkem je srozumitelnost dat. Je nezbytné, aby uložená data byla srozumitelná pro uživatele, kteří budou s daty pracovat. Srozumitelnost dat pro analytiku provádějící predikci CLV přispěje jednak k rychlejšímu provedení predikce, ale zamezí i případných chybám spojených se špatným porozuměním datům.

Důvěryhodnost dat zajišťuje ochotu využít data a důvěru ve správnost těchto dat. Data, která budou použita pro predikci CLV, musí být kvalitní. Nekvalitní data by ovlivnila přesnost odhadnuté CLV.

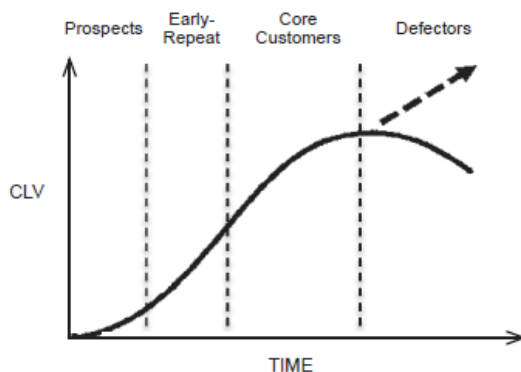
Výše uvedený souhrn požadavků na kvalitní data není konečný. Problémem kvality dat se zabývá mnoho literárních zdrojů. Dalším faktorem datové kvality je například dostupnost dat. Tento problém je tedy spíše cílen na zdroj dat pro predikci CLV. Je potřeba, aby ze CRM systému bylo možno snadno a rychle získat požadovaná data o zákaznících a mít možnost přistupovat k datům pro určité zvolené zákazníky.

Cílem této kapitoly bylo upozornit na tento problém, protože hraje významnou roli pro přesnost a kvalitu prováděných predikcí CLV. Podnik by měl věnovat zvýšenou pozornost kvalitě dat, čímž získá kvalitnější predikce CLV a větší výhody spojené s těmito predikcemi. Kromě technologického pohledu na kvalitu dat, je vhodné poukázat na business podstatu zákaznických dat. Tato data je značně obtížné získat za předpokladu, že predikci by neprováděli analytici samotného podniku. Obtížnost získání dat je spojena s tím, že tato data jsou pro podnik velmi důležitým zdrojem informací o jeho zákaznících a zdrojem získání konkurenční výhody. Právě proto by měl podnik věnovat kvalitě dat zvýšenou pozornost.

3.4. Životní cyklus zákazníka

Porozumění životnímu cyklu zákazníka je jedním z faktorů, který ovlivňuje predikce CLV. S jednotlivými fázemi životního cyklu zákazníka je spojeno rozdílné chování zákazníků, které je také ovlivňováno heterogenitou jednotlivých zákazníků. Na základě porozumění změnám v chování zákazníků lze zvýšit CLV zákazníků přizpůsobením různých marketingových strategií a kampaní. Na druhou stranu je však důležité si uvědomit, že životní cyklus také závisí na situaci, jde-li o contractual nebo non-contractual prostředí.

Rožek, Karlíček (2014) popisují situaci, kdy podnik využívá různé strategie pro jednotlivé fáze životního cyklu zákazníka. Obr. 4 zachycuje čtyři fáze životního cyklu v závislosti na CLV.



Obr. 4 - Životní cyklus zákazníka (Rožek, Karlíček 2014)

Každá z těchto fází životního cyklu určuje, o jaký typ zákazníka se jedná. Podnik proto musí přizpůsobit svůj přístup a vztah k zákazníkovi podle toho, v které fázi se nachází.

První fáze zahrnuje potencionální zákazníky, kteří se dosud rozhodují, jestli budou u podniku nakupovat. Nejde tedy zatím o zákazníky v tom pravém smyslu. Proto, aby podnik

získal tyto zákazníky s jejich potencionální hodnotou, musí využít strategie, která by zákazníka přilákala. Jedná se tedy o proces akvizice zákazníka, kdy podnik čelí problému jakým způsobem získat zákazníka pomocí různých výhod, aby nešel ke konkurenci.

V další fázi lze považovat tyto zákazníky za opravdové, jsou to ti, kteří již realizovali první transakci. Tito zákazníci se rozhodují o tom, zda jsou spokojeni s kvalitou nakoupeného zboží či služeb z první transakce. Na základě jejich pozitivního rozhodnutí lze očekávat, že budou realizovat další transakce. Podnik by se měl soustředit na kvalitu nabízeného zboží a služeb.

Třetí fáze je spojena s věrnými zákazníky, kteří nakupují pravidelně u daného podniku. Zde je nutné připomenout rozdíl mezi contractual a non-contractual prostředím. V první situaci lze mluvit o retenčním procesu, kdy podnik realizuje různé retenční kampaně, pomocí kterých se snaží udržet si své zákazníky co nejdéle, aby neodešli ke konkurenci. Druhá situace neřeší retenční proces, protože zákazníci mohou realizovat transakce i u jiných podniků. Do této fáze životního cyklu za této situace, lze zařadit zákazníky s vysokou frekvencí prováděných nákupů (viz kapitola 3.2.). Jedná se o zákazníky s nejvyšší hodnotou CLV, které si podnik snaží udržet co nejdéle, respektive se snaží podpořit, aby realizovali i nadále transakce s vysokou frekvencí opakování.

Následující fáze je spojena s odchodem zákazníků, kdy se zákazníci zamýšlejí nad odchodem nebo už dokonce nerealizují žádné další transakce. V contractual prostředí tedy zaniká vztah mezi zákazníkem a podnikem. Naopak v non-contractual prostředí jde o typ zákazníka C (viz kapitola 3.2.), kdy tento zákazník realizoval řadu transakcí, ale od poslední z nich už uplynula poměrně dlouhá doba. Podnik v této fázi řeší, zda je pro něj výhodné snažit se takového zákazníky udržet nebo je-li to příliš nákladné.

3.5. Hodnota zákazníka

Hodnotu zákazníků v jednotlivých fázích životního cyklu popisuje i Novo (2004). Na Obr. 5 lze pozorovat závislost současné a potencionální hodnoty zákazníka (dále jen CV a PV), které dohromady tvoří CLV zákazníka.

Current Value ↑	Low Potential Value, High Current Value Grow These Customers	High Potential Value, High Current Value Keep These Customers
	Low Potential Value, Low Current Value Should You Spend Money Here?	High Potential Value, Low Current Value Grow These Customers
Potential Value →		

Obr. 5 - Hodnota zákaznického portfolia (Novo, 2004)

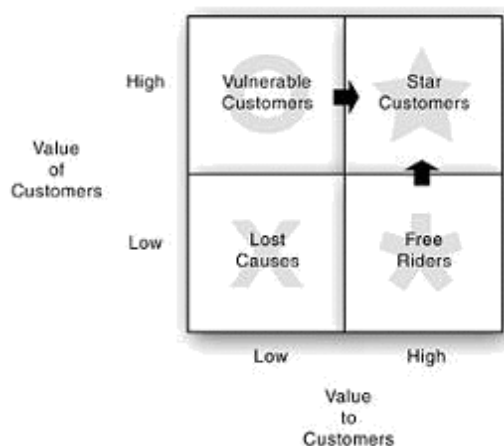
Levý dolní kvadrant představuje první fázi životního cyklu, kdy zákazníci mají nízkou CV i PV, a proto i jejich CLV je pro podnik nízká. Pravý dolní kvadrant popisuje druhou fázi životního cyklu, kdy zákazníci mají nízkou CV (dosud nerealizovali mnoho transakcí) a vysokou PV (lze očekávat jejich vysokou hodnotu v budoucnu). Třetí fázi životního cyklu popisuje horní pravý kvadrant, kdy zákazníci mají vysokou CLV. Horní levý kvadrant následně popisuje čtvrtou fázi životního cyklu. Tito zákazníci mají vysokou CV (realizovali mnoho transakcí), ale jejich PV je nízká, protože zaniká jejich vztah s podnikem, respektive v poslední době již nerealizují transakce.

Životní cyklus je tedy jedním z faktorů, který ovlivňuje CLV zákazníků. Je zásadní porozumět jednotlivým fázím, kdy se liší jak CV a PV zákazníka tak i jejich CLV. Je důležité, aby podnik správně vyhodnotil, v jaké fázi se zákazník nachází. Právě porozumění jednotlivým fázím a jakou CLV má zákazník, pomůže podniku při rozhodování jak přistupovat k jednotlivým skupinám zákazníků. Zásadní otázkou je jakou část rozpočtu a do které skupiny zákazníků by měl podnik investovat. Novo (2004) uvádí ve výše uvedené knize, že nejvíce profitabilní zákazníci jsou ti, kteří spadají do pravého horního kvadrantu a představují 80-90% zisku podniku, přičemž tvoří 10-20% zákaznické báze. Neznamená to však, že by měl podnik investovat jen do této skupiny. Důležité je vědět, jak k těmto ostatním skupinám zákazníků přistupovat, aby se z nich potencionálně mohli stát právě ti nejprofitabilnější zákazníci. Podnik by měl vidět v těchto zákaznících výzvu do budoucna a poučit se, proč nebyli zákazníci spokojeni s jeho produkty a nabízenými službami. Tuto situaci vystihuje citát B. Gatetese: „*Your most unhappy customers are your greatest source of learning.*“

Hodnota zákazníka a její vnímání

Kromě výše popsaného vztahu hodnoty zákazníka během životního cyklu, je třeba si uvědomit, že na hodnotu zákazníka lze nahlížet z různého pohledu. Gupta (2005) popisuje hodnotu zákazníka z dvou úhlů pohledu, které znázorňuje Obr. 6.

- Hodnota zákazníka pro podnik – zisk spojený se zákazníkem během jeho životního cyklu s podnikem
- Hodnota pro zákazníka – hodnotu, kterou podnik přináší zákazníkovi skrze nabízené produkty a služby



Obr. 6 - Hodnota zákazníka pro podnik a hodnota pro zákazníka (Gupta, 2005)

Vztah hodnoty zákazníka z dvou pohledů je charakterizován podobně jako vztah CV a PV (viz Obr. 5) pomocí čtyř kvadrantů, kde optimálním stavem je pravý horní kvadrant. Tento kvadrant představuje stav, kdy je hodnota zákazníka pro podnik v rovnováze s hodnotou pro zákazníka. Podnik poskytuje vysokou hodnotu pro zákazníka skrze nabídku kvalitních produktů a služeb. Zákazník zároveň přináší podniku vysokou hodnotu skrze zvýšení jeho loajality a delšího vztahu s podnikem. Opačným stavem je levý dolní kvadrant, kdy zákazník není spokojen s nabízenými produkty a službami, hodnota pro zákazníka i pro podnik je tedy nízká. Pravý dolní kvadrant představuje situaci, kdy podnik nabízí kvalitní produkty a služby, ale nezískává adekvátní hodnotu od zákazníka. Tato situace může nastat například z nátlaku konkurence, kdy se podnik nabízí řadu slev. Podnik se skrze slevy snaží přimět zákazníky k tomu, aby neodcházeli ke konkurenci a prodejem zlevněných produktů a služeb může v určité situaci docházet ke snižování hodnoty zákazníků pro podnik. Horní levý kvadrant naopak představuje situaci, kdy podnik získává od zákazníka více než zákazník od podniku. Příkladem mohou být zákazníci, kteří nově využívají produkty či služby podniku bez zvažování cenové politiky daného podniku, pro které by bylo výhodnější využívat případně levněji nabízené produkty či služby u konkurence. Proto tento vztah často vede ke ztrátě zákazníků.

Pro maximalizaci CLV v dlouhodobém časovém horizontu je optimální stav, kdy hodnota zákazníka pro podnik a hodnota pro zákazníka jsou v rovnováze. Spokojenost zákazníků s kvalitou nabízených produktů a služeb vede k dlouhodobému vztahu s podnikem, během kterého zákazníci přinášejí vysokou hodnotu pro podnik. Lze tak očekávat vyšší pravděpodobnost, že budou zákazníci aktivní i v budoucnosti a budou realizovat transakce s vyšší frekvencí. Jejich CLV bude tedy vyšší než ve zbylých třech případech.

3.6. Přínosy a výhody predikce CLV pro podnik

Využití predikce CLV přináší podniku řadu výhod. V předchozí kapitole bylo popsáno, jak CLV pomáhá podniku v rozhodování jakým způsobem přistupovat k jednotlivým skupinám zákazníků. Některé další přínosy jsou popsány například v článku Clark (2010).

První z výhod je spojena s predikcí CLV a získané CE. Jak již bylo zmíněno z CLV zákazníků lze získat CE, která slouží jako metrika pro kvantifikaci zákaznické báze podniku. Porovnáním hodnot CE za jednotlivé roky může podnik pozorovat vývoj hodnoty zákaznické báze. Podnik získá další ukazatel jeho ekonomického vývoje.

Pomocí CLV lze také zhodnotit přínosnost různých marketingových strategií. Například lze zhodnotit cílení podniku na rozdílné skupiny zákazníků, tedy jaké hodnoty CLV tyto skupiny přinesou. Dále lze sledovat vývoj nezměněné strategie získávání nových zákazníků a její dopad na CLV zákazníků. Podobně by šlo analyzovat strategii na udržení zákazníků a jednotlivých retenčních kampaní contractual prostředí. Právě toho využívají například české telekomunikační společnosti. Uvědomují si většího přínosu z retence zákazníků než, aby musely získávat nové zákazníky, jak popisuje Machander (2012).

Dále je možné na základě CLV zákazníků vytvořit jednotlivé segmenty s určitým rozmezím CLV. Pomocí této segmentace může podnik lépe alokovat rozpočet. V non-contractual prostředí je výhodné přidělit větší část na zákazníky, kteří mají větší pravděpodobnost, že jsou stále aktivními. Podnik může více uspořít, než kdyby při neznalosti CLV investoval větší část rozpočtu do segmentu zákazníků, kteří realizovali jen jednu transakci za dlouhou dobu a jsou pravděpodobně neaktivními.

Rovněž sledováním frekvencí nákupů a jejich odrazem v růstu či poklesu CLV, lze sledovat spokojenost zákazníků s kvalitou nabízených produktů a služeb.

Následně lze sledovat vliv různých komunikačních kanálů pro komunikaci podniku se zákazníkem. Jestli využití jednotlivých typů komunikace má vliv na zvýšení CLV a profitability zákazníka. Podnik pak může případně využívat od elektronické komunikace k osobní komunikaci pro jednotlivé segmenty zákazníků.

Pomocí CLV lze sledovat loajalitu zákazníků. Podnik může podpořit loajalitu zákazníků pomocí různých odměn ve formě slev a speciálních nabídek. Podnik tak může analyzovat vliv jednotlivých odměn a porovnat, které jsou výhodnější pro zvýšení CLV a profitability zákazníků.

Využitím predikce CLV zákazníků tedy podniku přináší plno výhod. Některé výhody jsou však spojeny se situací contractual prostředí a nemusí být výhodou pro non-contractual prostředí. Pro non-contractual prostředí by bylo zcestné zvažovat výhody plynoucí z retence zákazníků a zabývat se tím, zda zákazník odešel k jinému podniku. Přínosy predikce CLV jsou také úzce spjaté s kvalitou provedení těchto predikcí. Jak bylo zmíněno v předchozích kapitolách, kvalita predikce závisí mimo jiné na kvalitě zdroje dat a na kvalitě samotných dat. Přesto je důležité nahlížet na hodnoty CLV, jako na odhadovanou hodnotu. To je na jedné straně způsobeno těmito příčinami a na druhé straně je to důsledkem kvality použitých modelů a dalších proměnných, které podnik nedokáže ovlivnit.

4. Modely

Cílem této kapitoly je představit modely pro predikci CLV v non-contractual prostředí, které budou dále využity v praktické části práce. Vybranými modely jsou Pareto/NBD a BG/NBD a jsou využity v praktické části bakalářské práce pro predikce CLV zákazníků z analyzovaného datasetu. Tyto modely se dají využít pro non-contractual prostředí se spojitým rozdělením transakcí (viz kapitola 3.2).

4.1. Pareto/NBD model

Tento model slouží k predikci zákaznického chování, jako například učení pravděpodobnosti, že bude zákazník aktivní na konci určitého období a kolik realizuje transakcí. Pareto/NBD model patří mezi pravděpodobnostní modely aplikované pro non-contractual prostředí a byl vyvinut Schmittlein et al. (1987).

4.1.1. Pareto/NBD základní předpoklady a potřebná data

Přestože se tato bakalářská práce nezabývá do hloubky podstatou využitých modelů ze statistického hlediska (predikce jsou realizovány pomocí jazyka R), je vhodné představit základní předpoklady, ze kterých modely vycházejí. Pareto/NBD model vychází z šesti základních předpokladů. Tyto předpoklady popisuje Gupta, S. et al (2006) následovně:

- Vztah zákazníka s podnikem má dvě fáze: zákazník je aktivní po nepozorovatelnou dobu a potom se stává natrvalo neaktivním
- Po dobu kdy je zákazník aktivní, je počet transakcí realizovaných zákazníkem charakterizován Poissonovým rozdělením
- Heterogenita ve frekvenci transakcí mezi zákazníky je charakterizována gamma rozdělením
- Tzv. lifetime (celoživotní vztah zákazníka s podnikem) jednotlivého zákazníka je nepozorovatelný a je charakterizován exponenciálním rozdělením
- Heterogenita v drop out frekvenci (frekvence kdy zákazníci přestávají být aktivní) mezi zákazníky je charakterizována gamma rozdělením
- Frekvence transakcí a drop out se nezávisle liší mezi jednotlivými zákazníky

Tento model potřebuje relativně jen pár informací k tomu, aby mohl být použit k predikci zákaznického chování a následně CLV hodnoty. Vstupní data, z kterých Pareto/NBD model vychází, jsou následující:

1. Frekvence transakcí – informace o tom, kolik zákazník realizoval transakcí v daném období
2. Tzv. „recency“ – hodnota udávající, kdy zákazník realizoval poslední transakci

4.1.2. Parametry a získané informace z Pareto/NBD modelu

Pro aplikování Pareto/NBD modelu (rovněž BG/NBD) je nezbytné stanovit hodnoty jednotlivých parametrů daného modelu. Pareto/NBD je charakterizován čtyřmi parametry. První dva parametry r a α popisují gamma rozdělení heterogenity frekvence transakcí.

Následující dva parametry s a β popisují gamma rozdělení drop out frekvence. Odhad hodnot parametrů je dále realizován na analyzovaném datasetu v praktické části práce.

Jak již bylo zmíněno Pareto/NBD model slouží k predikci zákaznického chování a lze jeho pomocí realizovat predikce CLV. Podrobnější výstup informací získaných z tohoto modelu je proveden v praktické části práce. Proto je zde uvedeno několik základních informací, které model nabízí. Mezi hlavní výstupy modelu tedy patří:

- Očekávaný počet transakcí, které zákazník realizuje v testovacím období.
- Pravděpodobnost je-li zákazník aktivní v na konci kalibračního období.
- Závislost počtu aktivních zákazníků na konci kalibračního období a pravděpodobností, že jsou aktivní.
- Agregovaný odhad počtu transakcí realizovaných zákazníky v kalibračním a testovacím období.

Model nabízí různé charakteristiky zákaznického chování, které je možné porovnat i s daty obsaženými v datasetu. Tím lze porovnat, jak kvalitní výsledky daný model nabízí a aplikováním dalšího modelu (v této práci BG/NBD) za stejných podmínek lze porovnat modely mezi sebou.

4.2. BG/NBD model

Tento model byl vyvinut jako alternativní variantou pro Pareto/NBD model Fader (2005b). Částečně se liší v základních předpokladech pro zákaznické chování. Rozdíl je v tom, kdy se zákazník stává neaktivním. Pro Pareto/NBD se zákazník může stát neaktivním kdykoliv, zatímco u BG/NBD se stává zákazník neaktivním přímo po realizování transakce. Jedním z hlavních přínosů tohoto modelu je jeho snazší implementace.

4.2.1. BG/NBD základní předpoklady a potřebná data

Podobně jako Pareto/NBD vychází i BG/NBD model ze základních předpokladů. Platzer, M. (2008) vychází z článku Fader (2005), který definuje následující předpoklady:

- Po dobu kdy je zákazník aktivní, počet transakcí zákazníka je charakterizován Poissonovým rozdělením. Ekvivalentně doba mezi jednotlivými transakcemi (frekvence transakcí) je charakterizována exponenciálním rozdělením.
- Heterogenita ve frekvenci transakcí mezi zákazníky je potom charakterizována gamma rozdělením.
- Po každé transakci se stává zákazník neaktivním s určitou hodnotou pravděpodobnosti. Doba kdy se zákazník stává neaktivním (drop out) je charakterizována geometrickým rozdělením.
- Heterogenita v pravděpodobnosti, že se zákazník stane neaktivním je charakterizována beta rozdělením.
- Frekvence transakcí a frekvence drop out je variabilní nezávisle mezi zákazníky.

Předpoklady BG/NBD modelu jsou tedy dosti podobné jako předpoklady pro Pareto/NBD model. Stejná rozdělení pro počet transakcí a heterogenitu frekvence transakcí je stejné pro oba modely, což je důsledkem, že model vychází z historicky dřívějšího modelu NBD, jak popisuje ve své práci Platzer, M. (2008). Drobný rozdíl lze pozorovat u procesu, kdy se zákazník stává neaktivním. Pareto/NBD model vychází z předpokladu, že doba, kdy se stává zákazník neaktivním, je nepozorovatelná a nastává jednou. Zatímco BG/NBD model pracuje s předpokladem, kdy se zákazník stává s určitou pravděpodobností neaktivním po každé transakci. Proto tento model je charakterizována beta rozdělením pro hodnoty této pravděpodobnosti.

4.2.2. Parametry a získané informace z BG/NBD modelu

Tento model pracuje stejně jako Pareto/NBD se čtyřmi parametry. Tyto parametry jsou podobné jako u předchozího modelu r a α pro heterogenitu frekvence transakcí dále pak následují parametry a , b pro drop out.

Výstupy tohoto modelu jsou obdobné jako u Pareto/NBD. Odlišnost je možné sledovat v hodnotách výstupů, které model poskytuje. Právě srovnání predikovaných hodnot využitých modelů jsou srovnány v dále praktické části práce. Jak již bylo řečeno, vybrané paretové modely vycházejí z podobných předpokladů a pracují na podobném principu. Navíc BG/NBD model vznikl jako alternativou pro Pareto/NBD, a proto lze očekávat, že modely budou poskytovat podobné výsledky predikcí. Přesto jedním z cílů práce bude porovnat tyto výsledky a identifikovat jak a v jaké míře se liší.

4.3. Využití paretovských modelů k predikci CLV

Pro využití informací získaných z paretovských modelů k predikci CLV, je třeba tyto modely rozšířit o model, který vyjádří predikce počtu transakcí v peněžních částkách. Za tímto účelem je v této práci využit gamma-gamma spend model.

Fader (2005) popisuje využití tohoto modelu a důvod proč je potřeba využít modelu pro predikce průměrné hodnoty transakce. Přestože jsou obvykle známy tržby za jednotlivé transakce, není obtížné stanovit i průměrnou hodnotu tržeb pro jednotlivé zákazníky. Problém je však v tom, že zákazníci, kteří realizovali jen jednu transakci, nemusí poskytovat přesnou průměrnou hodnotu transakcí. Zákazník mohl například nakoupit určité zboží či služby za nízkou hodnotu, ale v budoucnosti může realizovat další transakce. Může následně realizovat transakce za vyšší hodnoty a dokonce i daleko častěji. Tím skutečná průměrná cena transakcí tohoto zákazníka vzroste. Využití gamma-gamma modelu tak nabízí možnost, jak odhadnout průměrnou hodnotu transakcí.

4.3.1. Gamma-gamma model základní předpoklady

Tento model vychází ze základních předpokladů, které popisuje Fader (2005).

- Hodnoty transakcí jednotlivých zákazníků se náhodně liší od jejich skutečné nepozorovatelné průměrné hodnoty

- Průměrná hodnota transakcí je odlišná mezi jednotlivými zákazníky, ale tato hodnota se neliší v čase.

Ze základních předpokladů tedy vyplývá, že každý pro každého zákazníka existuje určitá průměrná hodnota jeho transakcí. Tato hodnota však není pozorovatelná, s větším počtem transakcí realizovaných zákazníkem je možné se přiblížit k jeho skutečné průměrné hodnotě.

Parametry a získané informace z modelu

Gamma-gamma model pracuje se třemi parametry (p , q , γ). K odhadu těchto parametrů jsou zapotřebí frekvence transakcí zákazníků a průměrné hodnoty transakce jednotlivých zákazníků. Pomocí gamma-gamma modelu lze následně získat odhadované průměrné hodnoty transakcí pro jednotlivé zákazníky, které lze využít zejména pro budoucí odhad, kdy nejsou již dostupná data zákazníků.

4.3.2. Výpočet CLV

Pomocí výstupů získaných z paretovských modelů a gamma-gamma modelu lze již vypočítat CLV hodnoty zákazníků. Predikce CLV budou realizovány pomocí základního vzorce, který uvádí (Gupta, S. et al, 2006).

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

kde

p_t - je peněžní hodnota transakce placené zákazníkem

c_t - jsou náklady přímo spojené s transakcí

i - je diskontní míra

r_t - je pravděpodobnost, že bude zákazník aktivní v čase t

AC - jsou náklady na získání zákazníka

T - je časový horizont, pro který je predikována CLV hodnota

Pro účely bakalářské práce bude vyjádření jednotlivých nákladů zjednodušeno. Jednotlivé peněžní hodnoty transakcí budou sníženy na 15 procent. Toto snížení právě představuje jednotlivé náklady. Predikce CLV budou realizovány pro časový horizont tří let a diskontní míra je zvolena na dvě procenta. Predikce CLV hodnot v praktické části bakalářské práce budou vypočítávány pomocí predikovaného počtu transakcí pro jednotlivá období vyjádřených v peněžní hodnotě s využitím průměrné hodnoty transakce jednotlivých zákazníků získané z gamma-gamma modelu. Tato hodnota je pro jednotlivá období diskontována s využitím zvolené diskontní míry dvou procent a vynásobena pravděpodobností, že bude zákazník v daném období aktivní.

5. Praktická část – aplikování metodiky CRISP-DM

Cílem kapitoly je využít zavedenou metodiku CRISP-DM pro práci s datasetem internetového obchodu. V této kapitole tak budou aplikovány jednotlivé fáze metodiky. Bude zde řešeno porozumění businessu, porozumění datům, přípravě dat, modelování (kalibrace modelů a odhad parametrů) a zhodnocení výsledků modelů. Využití výsledků bude diskutováno v kapitole 6.4.

Pro praktickou část práce je využít statistický nástroj R od R Core Team (2015). Prováděné analýzy jsou realizovány v softwarovém prostředí nástroje RStudio verze 0.98.1103 s programovacím jazykem R verze 3.1.3 (dále jen R). Pro implementaci Pareto/NBD a BG/NBD modelů jsou využity funkce z balíčku BTYD od Dziurzynski (2014).

5.1. Fáze porozumění businessu

Podle metodiky CRISP-DM je nezbytné porozumět v první fázi podniku a jeho business záležitostem pro porozumění samotným datům. Analyzovaný dataset pochází z reálného internetového obchodu. Pro zachování jeho anonymity není v této práci blíže určeno, o jaký internetový obchod se jedná ani jaké produkty nabízí. Pro porozumění zákaznickému chování je třeba se podívat na frekvence realizovaných objednávek jednotlivých zákazníků.

Za období 17. 10. 2010 až 29. 3. 2015 nakupovalo na internetovém obchodu 35 249 zákazníků, z nichž 13 226 nakupovalo více než jednou. Na některé z těchto zákazníků připadají stornované objednávky, a proto bude počet analyzovaných zákazníků nižší po odstranění zrušených objednávek (viz kap. 5.3). Za celé sledované období 232 týdnů lze sledovat rostoucí trend počtu objednávek. Z celkového počtu 48 475 objednávek jich bylo realizováno 17 680 v prvních 116 týdnech, zatímco v následující polovině sledovaného období jich bylo realizováno již 30 795.

Základní hodnoty pro množství realizovaných objednávek zákazníky jsou následující:

- Minimální počet objednávek: 1 objednávka (každý zákazník realizoval alespoň 1 objednávku)
- Maximální počet objednávek: 64 objednávek
- Medián objednávek: 1 objednávka
- Průměrný počet objednávek: 1,375 objednávek

Tabulka 1 obsahuje podrobnější informace o realizaci objednávek pro vybrané zákazníky. Některé objednávky jsou však stornované, a proto u některých zákazníků po fázi přípravy dat bude frekvence objednávek nižší.

Tabulka 1 - Množství objednávek pro vybrané zákazníky (Zdroj: autor)

Identifikátor zákazníka	Objednávka	1. - 116. týden	117. - 232. týden	Storno objednávky
264	64	55 (1 Storno)	9	1
167	29	26	3	0
1	1	0	1	0

Zákazník 264 realizoval nejvíce objednávek a v období prvních 116 týdnů naposledy realizoval objednávku 2. 1. 2013 (55 objednávek za 808 dní). Zákazník 167 realizoval 26 objednávek ve stejném období a poslední z nich 28. 11. 2012 (26 objednávek za 773 dní). Zákazník 264 tak dosahuje zhruba dvojnásobné frekvence. Podle historie transakcí lze očekávat, že tento zákazník bude realizovat více objednávek i v dalším období a také má větší pravděpodobnost být aktivní (viz Obr. 1). Zákazník 1 na druhé straně realizoval jen jednu objednávku a nelze proto očekávat, že by byl profitabilní. U jednotlivých zákazníků internetového obchodu tak lze sledovat rozdílné chování. Poměrně velká část zákazníků realizovala pouze jednu objednávku. Toto chování není neobvyklé pro non-contractual prostředí, lze tedy očekávat, že někteří z nich budou aktivní i za pár let.

5.2. Porození datů

Analýzovaný dataset zahrnuje údaje z jednotlivých objednávek internetového obchodu. Pro zachování anonymity dat má zkoumaný dataset data v anonymizované podobě. Obsahuje zahashované ID zákazníků a transakcí (customerId a transactionId). Pro lepší práci s customerId a názornější interpretaci individuálních predikcí bude vhodné jeho „hash“ hodnotu nahradit za hodnoty v číselné podobě.

Dataset obsahuje 48 475 transakcí (objednávek) za období 17. 10. 2010 až 29. 3. 2015, které jsou identifikovány devíti atributy. Každá transakce zahrnuje Id transakce a zákazníka v zahashované podobě a sedm dalších atributů (viz Tabulka 2). V datasetu nejsou zahrnuty duplicitní objednávky ani se v něm nevyskytují nevyplněné hodnoty atributů. Ve fázi přípravy dat tedy není nutné řešit problém neúplných dat.

Tabulka 2 - Charakteristika datasetu (Zdroj: autor)

Atributy	Minimum	Medián	Průměr	Maximum
date	17. 10. 2010	5. 9. 2013	23. 3. 2013	29. 3. 2015
transactions	0	1	1	3
transactionRevenue	-1525 Kč	777 Kč	1049 Kč	56970 Kč
transactionShipping	-79 Kč	69 Kč	56	299
transactionTax	-631,63 Kč	122,81 Kč	172,08 Kč	9495 Kč
itemQuantity	-7	2	2,575	72
Status	Vyřízená		Storno	
	47332		1143	

Popis atributů:

- transactions - Udává, zda objednávka byla upravována. Hodnoty jiné než 1 označují objednávky, které byly nejspíše upravovány. Jedná se o zanedbatelné množství objednávek (celkem 111 upravovaných objednávek).
- transactionRevenue - Představuje tržby z objednávek.
- transactionShipping - Udává částku za dopravu.
- transactionTax - DPH z objednávky odvozené z výše transactionRevenue.
- itemQuantity - Počet kusů zboží v objednávce.
- Status - Nabývá dvou hodnot: „Vyřízená“ pro úspěšné vyřízení objednávky nebo „Storno“ pro zrušené objednávky.

Z hodnot jednotlivých atributů je patrné, že před samotným využitím dat pro predikci CLV bude nutné věnovat se také třetí fázi metodiky CRISP-DM - přípravě dat. Bude nezbytné odstranit zrušené objednávky a ošetřit záporné hodnoty atributů apod.

Pro využití Pareto/NBD a BG/NBD modelů jsou nejdůležitějšími daty customerId, transactionRevenue a date. Atribut CustomerId identifikuje zákazníka, který realizoval transakce a pomocí atributu date budou získány informace o frekvenci transakcí a tom, kdy zákazníci realizovali poslední transakci. Následné predikce CLV abstrahují od výpočtu čisté hodnoty po zdanění, proto atribut transactionTax nebude v predikcích zahrnut. Náklady spojené s jednotlivými transakcemi budou zjednodušeny tím způsobem, že jednotlivé tržby budou sníženy na patnáct procent z původní hodnoty. Toto procentuální snížení bude představovat snížení transactionRevenue o příslušné náklady a DPH. Proto atribut transactionShipping nebude rovněž využit.

5.3. Příprava dat

Důležitou fází definovaná metodikou CRISP-DM je příprava dat. Před samotnou analýzou dat je potřeba připravit data v požadované formě, která vyplývá z dané analýzy a dále také zabránit provedení analýzy na nekvalitních datech (viz kapitola 3.3.2). Data je tak nutno přizpůsobit požadavkům softwarového prostředí, v němž je analýza prováděna. Na druhé straně ne všechna data jsou nezbytná, a proto je během přípravy dat kladen nárok na výběr relevantních dat pro prováděnou analýzu.

Dataset internetového obchodu obsahuje pro jednotlivé transakce všechna data. Nebylo tak nutné řešit problém neúplnosti dat pro jednotlivé zákazníky. Na druhou stranu v datasetu jsou obsažena některá data, která pro predikci CLV pomocí paretovských modelů s využitím jazyka R nejsou potřebná. Dále dataset obsahuje i data, jenž je nutné před provedením analýzy vyloučit.

Z předchozí fáze věnované porozuměním datům vyplývají následující požadavky na přípravu dat.

- Odstranění zrušených objednávek
- Ošetření záporných a nulových hodnot
- Převod `customerId` na numerickou hodnotu
- Oddělovač desetinných míst
- Datum v požadovaném formátu
- Sloučení transakcí

Odstranění zrušených objednávek

Odstranění zrušených objednávek zahrnuje objednávky se „Storno“ hodnotou atributu „Status“, které nebyly realizovány. Z datasetu bylo tak odstraněno 1 143 zrušených objednávek, které realizovalo 1 039 zákazníků.

Ošetření záporných a nulových hodnot

S předchozí částí přípravy dat souvisí i požadavek na ošetření záporných a nulových hodnot. Většina nežádoucích záporných hodnot (např. záporný počet kusů zboží a další) byla odstraněna spolu se stornovanými objednávkami, mohlo se jednat například o špatně vyplněné objednávky. Přesto zde zůstalo 488 objednávek se záporným DPH (atribut `transactionTax`). Při využití tohoto atributu pro predikce CLV by bylo nutné, ošetřit tyto záporné hodnoty. Dále bylo potřeba odstranit 28 objednávek s nulovou hodnotou `transactionRevenue`, aby tak analyzované objednávky nezahrnovaly i objednávky s touto hodnotou.

Po odstranění storno objednávek, záporných a nulových hodnot je získán výsledný počet 34 631 zákazníků. Pro tyto zákazníky bude možné predikovat jejich CLV hodnotu (viz kap. 6.).

Převod `customerId` na číselnou hodnotu

Jednotlivá `Id` zákazníků jsou převedena na numerickou hodnotu. Prvnímu zákazníkovi v datasetu bylo přiřazeno `Id` s hodnotou 1. Pro dále uvedené zákazníky je hodnota jejich `Id` zvětšena o 1 oproti zákazníkovi předcházejícímu. Jednotlivé objednávky realizované stejným zákazníkem jsou obdobně identifikovány unikátním numerickým `Id` zákazníka.

Následující části přípravy dat již souvisí požadavky parietovských modelů a softwarového prostředí jazyka R. Pro úplnost lze uvést, že jazyk R pracuje s tečkou pro oddělování desetinných míst. Jsou-li v datasetu oddělována desetinná místa čárkou, je nutné určit vstupním parametrem funkce pro načítání csv, že je ve zdrojovém souboru využita desetinná čárka. Autoři walkthrough k balíčku `Buy 'Til You Die` (dále jen `BTYD`) McCarthy, Wadsworth (2014) popisují mimo jiného právě přípravu dat pro využití Pareto/NBD a

BG/NBD modelu. Je zde popsána část přípravy dat věnující se formátu data, výběru potřebných dat a sloučení transakcí. Následuje popis těchto úprav.

Datum a jeho formát

Data je třeba transformovat do jednotného formátu, aby při jejich porovnávání nedocházelo ke špatné interpretaci mezi jednotlivými formáty. Data z původního datasetu byla upravena do formátu „rok-měsíc-den“. Pomocí této úpravy lze v R následně chronologicky řadit jednotlivé objednávky.

Sloučení transakcí

Jelikož oba modely Pareto/NBD i BG/NBD pracují s časem mezi jednotlivými objednávkami, je třeba provést sloučení určitých transakcí. U objednávek ze stejného dne nelze již určit dobu mezi jejich realizací, proto je nutné sloučit tyto objednávky. Pomocí funkce „dc.MergeTransactionsOnSameDate“ z balíčku BTYD bylo provedeno sloučení objednávek realizovaných ve stejný den. Dataset obsahuje 501 objednávek ze stejného dne, sloučeno bylo 253 z nich (zbylých 253 bylo vyloučeno v rámci vyloučení stornovaných objednávek).

5.4. Kalibrace modelů a odhad parametrů

Tato kapitola představuje další fázi metodiky CRISP-DM. V této fázi je potřeba provést kalibraci modelů. Pro paretovské modely je nutné provést odhad parametrů, s jejichž využitím bude následně možno využít modely k predikcím zákaznického chování a CLV. Cílem kapitoly je tedy na analyzovaném datasetu natrénovat modely, aby následně podávaly co nejpřesnější výsledky.

Odhad parametrů

Pro odhad parametrů je třeba nejprve rozdělit dataset na tzv. kalibrační a testovací (validační) období. Kalibrační období tak představuje část dat, na kterých je proveden odhad parametrů. Se získanými parametry je již možné využívat modely k predikci zákaznického chování. Data z testovacího období následně slouží ke zhodnocení přesnosti modelu. Lze volit různá procentuální rozdělení datasetu pro tato období. Čím delší je zvoleno kalibrační období, tím lze očekávat přesnější odhad parametrů. Na druhé straně přesnost modelu je následně možno realizovat na menším vzorku dat.

Odhad parametrů je realizován v R s využitím funkcí z balíčku BTYD. Poměr rozdělení datatasetu na kalibrační a testovací období je v této práci zvolen 50/50 a 90/10. Tabulka 3 obsahuje získané hodnoty parametrů společně s metrikami zákaznického chování určenými poměrem těchto parametrů.

Tabulka 3 - Odhad parametrů (Zdroj: autor - realizováno v R)

Pareto/NBD	r	α	r/α	s	β	s/β
Rozdělení 50/50	0,6849	102,8123	0,0067	0,1712	2,1729	0,0788
Rozdělení 90/10	0,8226	134,9261	0,0061	0,4628	27,3258	0,0169
BG/NBD	r	α	r/α	a	b	$a/a+b$
Rozdělení 50/50	0,1999	33,1338	0,006	1,5209	2,3574	0,3922
Rozdělení 90/10	0,2163	34,2956	0,0063	2,3895	3,7039	0,3921

Význam metrik popisuje Wübben (2008), kde r/α pro Pareto/NBD model udává počet transakcí, který realizuje průměrný zákazník za určitou časovou jednotku (v případě této bakalářské práce jsou časovou jednotkou týdny). Dále s/β vyjadřuje drop out frekvenci pro průměrného zákazníka za časovou jednotku. S využitím metriky s/β lze následně určit po jakou dobu zůstane průměrný zákazník aktivní.

Tato doba je určena poměrem: $\frac{1}{s/\beta}$

Pro rozdělení 50/50 tak vychází, že bude zákazník aktivní 12,69 týdnů a v případě rozdělení 90/10 vychází tato doba na 59,17 týdnů.

Podobně pro BG/NBD model r/α udává počet transakcí průměrného zákazníka za časovou jednotku. Doba, po kterou zůstává zákazník aktivní, je vyjádřena v případě tohoto modelu poměrem: $\frac{1}{\left(\frac{r}{\alpha}\right) * \left(\frac{a}{a+b}\right)}$

Tato hodnota vychází 404,33 týdnů pro rozdělení 50/50 a pro rozdělení 90/10 následně 404,82 týdnů, po které bude zákazník aktivní.

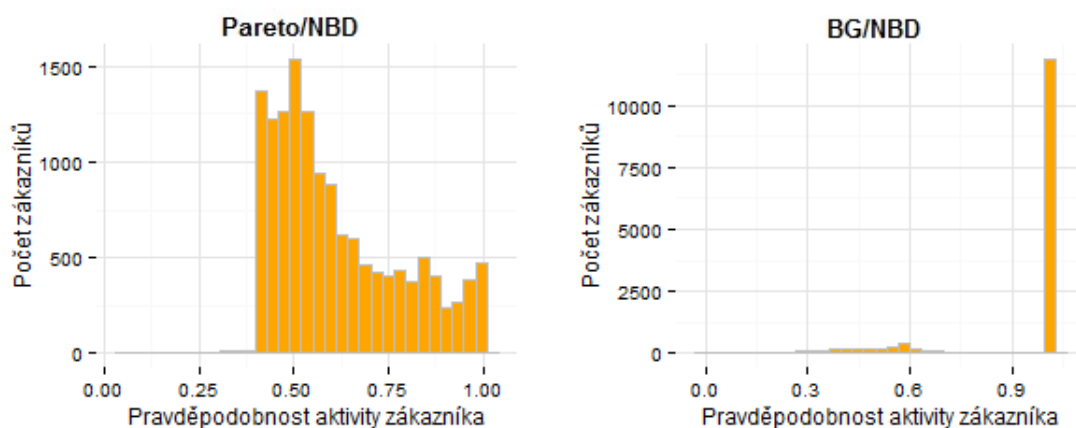
5.5. Zhodnocení modelů

V této kapitole jsou již prezentovány výsledky jednotlivých modelů a jejich srovnání s reálnými daty. S využitím získaných parametrů lze pomocí paretovských modelů predikovat zákaznické chování. Lze odhadnout pravděpodobnost, zda jsou zákazníci aktivní na konci kalibračního období. Dalším důležitým výstupem modelů jsou predikce počtu transakcí zákazníků. V následující kapitole jsou provedeny predikce počtu transakcí pomocí Pareto/NBD a BG/NBD. Dále již následuje zhodnocení gamma-gamma spend modelu, jehož výstupem jsou peněžní hodnoty jednotlivých transakcí.

5.5.1. Pravděpodobnost aktivních zákazníků na konci kalibračního období

Jednou ze základních manažerských otázek je určit, zda bude zákazník aktivní v určitém období. Využitím paretovských modelů lze predikovat, bude-li zákazník aktivní na konci kalibračního období. Pro jednotlivá rozdělení datasetu bude délka tohoto období rozdílná.

Na Obr. 7 jsou zachyceny predikce paretovských modelů pro rozdělení datasetu, v kterém toto období tvoří 50% celkového sledovaného období (116 týdnů).



Obr. 7 - Pravděpodobnost aktivity zákazníků na konci kalibračního období (Zdroj: autor)

Kalibrační období s délkou 116 týdnů zahrnuje objednávky, které realizovalo 14 087 zákazníků. Výsledky těchto modelů, jsou na první pohled velmi odlišné. Výsledky BG/NBD modelu predikují téměř pro všechny zákazníky skoro sto procentní pravděpodobnost, že budou aktivní. Příčinu vysoké frekvence zákazníků s vysokou pravděpodobností popisuje Fader (2008). Zákazníci, kteří nerealizovali žádnou objednávku, mají pravděpodobnost rovnou 1. Tato vlastnost vyplývá ze základních předpokladů BG/NBD modelu, který předpokládá, že každý zákazník je na začátku sledovaného období aktivní. Pro rozdělení datasetu 50/50 zahrnuje kalibrační období 11 879 zákazníků, kteří realizovali jen jednu objednávku. Protože paretové modely zkoumají opakované objednávky, dochází u těchto zákazníků ke snížení jejich frekvence počtu transakcí na nulovou hodnotu. Právě proto lze pro tyto zákazníky sledovat vysokou hodnotu pravděpodobnosti, že budou aktivní na konci kalibračního období (Obr. 7).

5.5.2. Srovnání transakcí pro kalibrační a testovací období

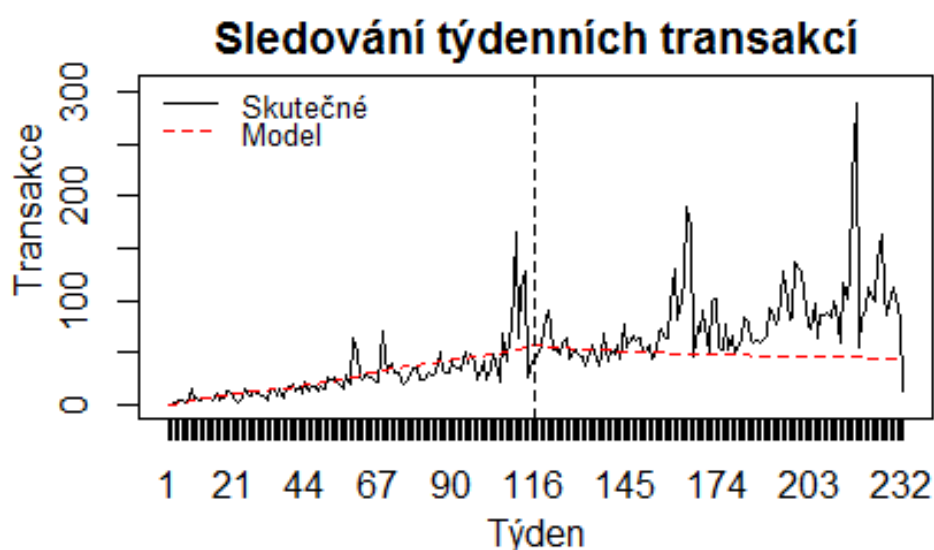
Dalším z hlavních výstupů Pareto/NBD a BG/NBD modelů jsou predikce očekávaného počtu transakcí, které zákazníci budou realizovat. Modely tak nabízejí odpověď na to, kolik lze v určitém období očekávat transakcí realizovaných zákazníky. Lze tak provádět individuální predikce pro jednotlivé zákazníky nebo agregované predikce pro všechny zákazníky. Pro kvalitu a přesnost těchto predikcí je nezbytné provést co nejpřesnější odhad parametrů a následně modely otestovat na reálných datech. Srovnání výsledků modelů s reálnými daty poskytuje možnost, jak zhodnotit kvalitu predikcí výsledky modelů mezi sebou. Ze získané analýzy kvality modelů lze následně určit, který model je přesnější a do jaké míry se lze spolehnout na jeho výsledky. Jak již bylo zmíněno, CLV je hodnota, jejíž význam je spjat s budoucností. Proto je hlavním významem predikcí modelů určit její hodnotu od současnosti do budoucna.

K zhodnocení Pareto/NBD a BG/NBD modelů je využit dataset internetového obchodu. Tento dataset byl pro predikce očekávaného počtu transakcí rozdělen na kalibrační a testovací období v poměru 50/50 a také 90/10. Rozdělení v poměru 50/50 tedy poskytuje polovinu dat pro kalibraci modelů a druhou polovinu k otestování modelů. Druhá varianta

rozdělení datasetu potom poskytuje více dat pro přesnější kalibraci modelů, ale následně je k dispozici méně dat ke zhodnocení.

Následující predikce jsou inspirovány podle Baggott, M. (2013), McCarthy, D., Wadsworth, E. (2014) a Platzer, M. (2008) jejichž realizace proběhla v R. Délka období datasetu je 232 týdnů. Na níže uvedených grafech lze sledovat agregovaný odhad počtu objednávek v jednotlivých týdnech ve srovnání se skutečně realizovaným množstvím.

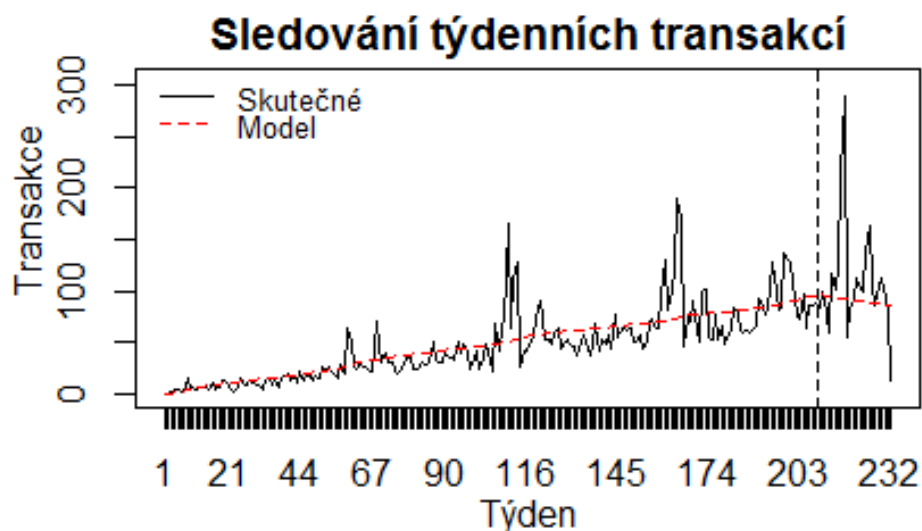
Na Obr. 8 jsou zachyceny výsledky Pareto/NBD modelu pro rozdělení datasetu v poměru 50/50. Na reálných datech lze sledovat plynulý růst počtu objednávek (transakcí) od prvního až do posledního týdne s několika výraznými výkyvy, kdy bylo realizováno více objednávek v porovnání s celkovým trendem.



Obr. 8 - Pareto/NBD s rozdělením 50/50 (Zdroj: autor)

Dále je možné sledovat, že odhady modelu v testovacím období nejsou zrovna přesné v porovnání s reálnými daty a model tak podceňuje reálný počet objednávek. Zcela nejvyšší rozdíl mezi odhadem modelu a reálnými daty je vidět v 217. týdnu, kdy bylo realizováno 288 objednávek. Naopak model nadhodnocuje počet objednávek například hned pro 116. týden a nejvíce pro poslední týden.

Následující graf na Obr. 9 zachycuje rozdělení v datasetu pro Pareto/NBD model v poměru 90/10. Na první pohled lze pozorovat přesnější výsledky modelu pro testovací období v porovnání s předchozím rozdělením, nicméně toho zlepšení lze sledovat jen na jedné desetině dat. Kalibrace modelu na větším počtu dat vedla tedy k zvýšení kvality modelu pro poslední desetinu pozorovaného období.



Obr. 9 - Pareto/NBD s rozdělením 90/10 (Zdroj: autor)

Tabulka 4 prezentuje skutečná a predikovaná množství objednávek pomocí Pareto/NBD modelu natrénovaného podle jednotlivých rozdělení datasetu. Ve sloupcích jsou zahrnuta množství objednávek v jednotlivých týdnech a dále celková množství za určité časové intervaly.

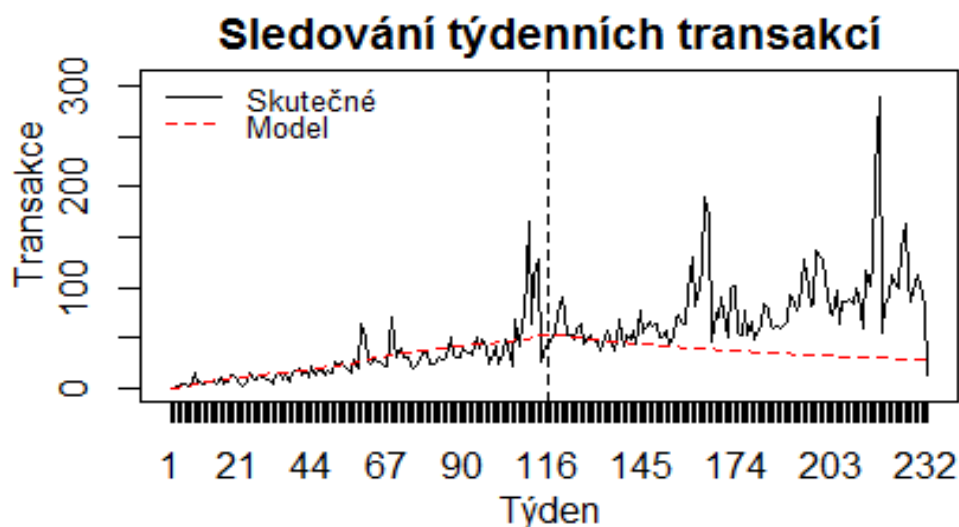
Tabulka 4 - Predikce počtu transakcí Pareto/NBD modelem (Zdroj: autor - realizováno v R)

Týden	1.	116.	232.	1. - 116.	117. - 232.	1. - 208,8.	208,9. - 232.	Přesnost 1. - 232.
Skutečný počet obj.	1	42	14	3149	9811	9288	2700	-----
Odhad Pareto/NBD 50/50	0,23	57,22	44,68	3203,18	7796,41	5679,29	1087,31	-28,54%
Odhad Pareto /NBD 90/10	0,21	55,78	85,70	3216,22	10003,43	8965,68	2188,68	-1,96%

Z vybraných týdnů lze pozorovat, že Pareto/NBD model v obou variantách jeho kalibrace nadhodnocuje i podhodnocuje skutečný počet objednávek. První varianta modelu pro tyto vybrané týdny poskytuje o něco bližší výsledky ke skutečným množstvím objednávek. Pro prvních 116 týdnů je již rozdíl nepatrný. Na druhé straně druhá varianta kalibrace modelu poskytuje přesnější výsledky, které celkově podhodnocují reálná data jen o 1,96 %.

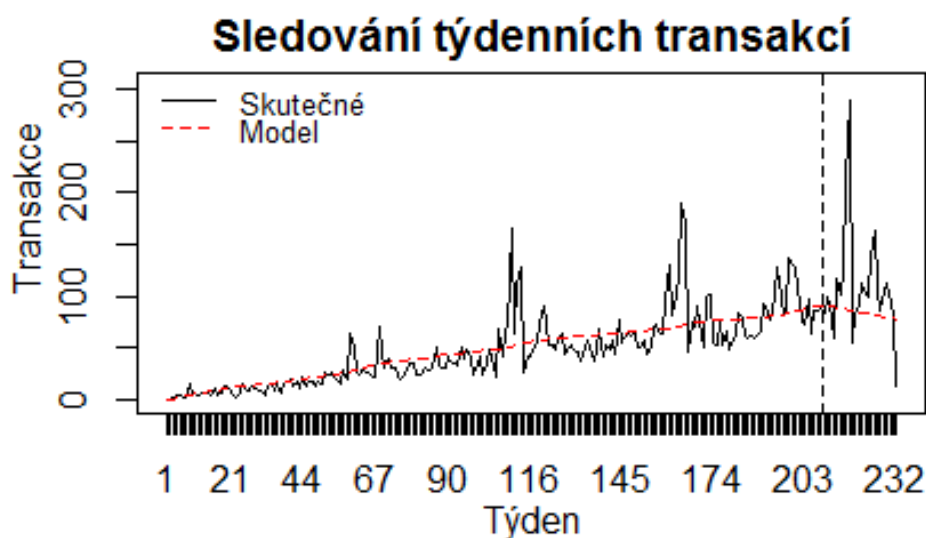
Následující graf na Obr. 10 již představuje predikce očekávaného množství transakcí pomocí BG/NBD modelu. Tento model byl rovněž kalibrován na stejných variantách rozdělení datasetu. Na první pohled lze pozorovat podobné výsledky tohoto modelu.

Přesto v testovacím období poskytuje ještě více podhodnocené predikce množství objednávek.



Obr. 10 - BG/NBD s rozdělením 50/50 (Zdroj: autor)

Na Obr. 11 jsou zobrazeny výsledky BG/NBD modelu pro rozdělení datasetu v poměru 90/10. Využití více dat ke kalibraci rovněž vedlo k zlepšení kvality modelu pro poslední desetinu dat.



Obr. 11 - BG/NBD s rozdělením 90/10 (Zdroj: autor)

Tabulka 5 nabízí stejná srovnání výsledků jako u Pareto/NBD modelu. Z výsledků BG/NBD modelu vyplývá, že tento model více podhodnocuje skutečný počet objednávek než Pareto/NBD model.

Tabulka 5 - Predikce počtu transakcí BG/NBD modelem (Zdroj: autor - realizováno v R)

Týden	1.	116.	232.	1. - 116.	117. - 232.	1. - 208,8.	208,9. - 232.	Přesnost 1. - 232.
Skutečný počet obj.	1	42	14	3149	9811	9288	2700	-----
Odhad BG/NBD 50/50	0,21	54,05	28,70	3166,61	6905,95	4460,79	724,39	-38,65%
Odhad BG/NBD 90/10	0,22	55,90	78,16	3290,73	9875,93	8615,73	2043,50	-4,18%

Ze sloupce přesnosti lze pozorovat, že při rozdělení 90/10 již tento model poskytuje poměrně přesné výsledky. Pro kratší období predikcí tak lze očekávat, že i tento model bude poskytovat v celku přesné výsledky.

Zhodnocení predikcí pomocí a RMSE a MSLE

K zhodnocení přesnosti predikovaného počtu objednávek s reálnými daty lze využít RMSE (root mean squared error) a MSLE (mean squared logarithmic error). Aplikováním těchto „metrik“ pro oba paretové modely lze získat odchylky predikcí od reálných dat. Tím lze získat porovnání modelů a stanovit, který z modelů je přesnější.

Tabulka 6 obsahuje srovnání Pareto/NBD a BG/NBD modelu pomocí RMSE a MSLE.

Tabulka 6 - Zhodnocení modelů RMSE a MSLE (Zdroj: autor - realizováno v R)

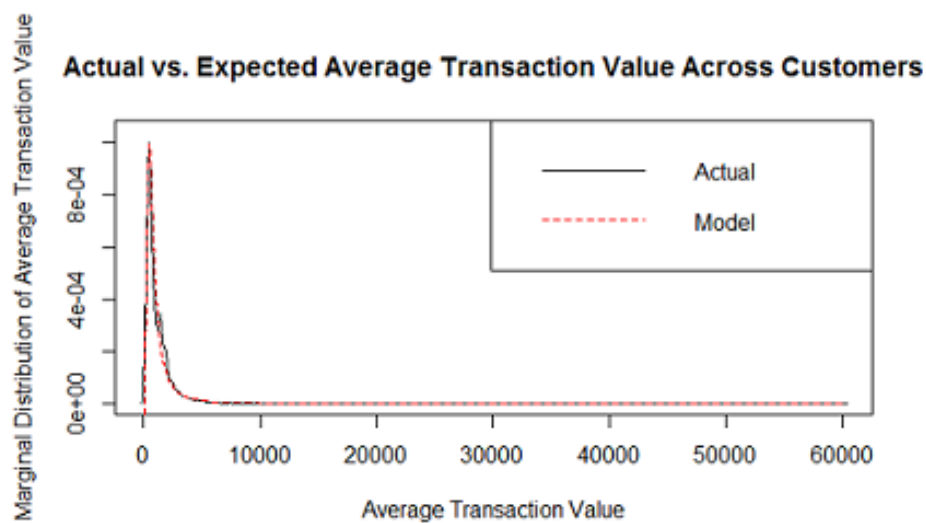
Model / rozdělení	RMSE	MSLE
Pareto/NBD 50/50	0,7769	0,1612
Pareto/NBD 90/10	0,2943	0,0347
BG/NBD 50/50	0,7515	0,1418
BG/NBD 90/10	0,294	0,0344

Pareto/NBD modely nabízejí v celku srovnatelné výsledky. O něco menší odchylky predikcí poskytuje BG/NBD model, ale pro rozdělení 90/10 jsou již výsledky modelů téměř stejné. Predikce modelů lze považovat za srovnatelné.

5.5.3. Gamma-gamma model

Jak již bylo řečeno v kapitole 4.3., paretové modely je potřeba rozšířit o submodel, kterým lze vyjádřit CLV v peněžní hodnotě. V této práci je využit tzv. gamma-gamma spend model, který nabízí odhadované průměrné hodnoty transakcí. Využití modelu poskytuje možnost lépe vyjádřit průměrnou hodnotu objednávek. Využití tohoto modelu je vhodné zejména pro predikce do budoucnosti, kdy podnik nemá již data o zákaznících.

Na Obr. 12 je vidět srovnání odhadovaných průměrných hodnot objednávek pomocí gamma-gamma modelu.



Obr. 12 - Gamma-gamma model zhodnocení predikcí (Zdroj: autor)

Je patrné, že tento model pro analyzovaný dataset nabízí přesné odhady průměrných hodnot objednávek. Proto ho lze využít pro predikce CLV v peněžní hodnotě v následující kapitole 6.

6. Predikce CLV

Tato kapitola zahrnuje predikce CLV pro zákazníky z analyzovaného datasetu internetového obchodu. K predikci jsou využity Pareto/NBD a BG/NBD modely s rozšířením o gamma-gamma model pro vyjádření peněžní hodnoty predikovaných počtů transakcí z paretovských modelů. Tato kapitola obsahuje individuální predikce CLV a dále predikce pro vyjádření CE. Závěr kapitoly poskytuje srovnání výsledků jednotlivých modelů.

Jednotlivé predikce jsou pro období následujících tří let po sledovaném období (po 232. týdnu). Pro dosažení co nejvyšší přesnosti predikcí, jsou využita data za všech 232. týdnů. Jednotlivé objednávky za toto období byly realizovány 34 631 zákazníky, pro které tak lze predikovat jejich CLV hodnotu.

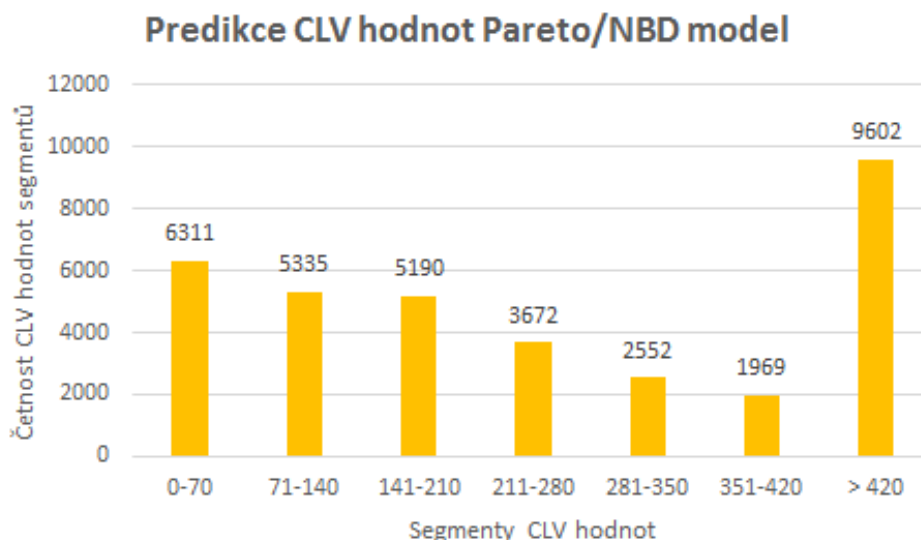
6.1. Individuální predikce

Individuální predikce CLV nabízí podniku možnost jak určit, jakou přinesou jednotliví zákazníci hodnotu pro podnik do budoucna. Na základě této informace se může podnik lépe rozhodovat o tom, kteří zákazníci jsou profitabilní. Lze přizpůsobit výši investic podle CLV hodnot jednotlivých zákazníků a přizpůsobit jim marketingové plány a kampaně.

6.1.1. Predikce CLV pomocí Pareto/NBD modelu

Pro predikce CLV hodnot zákazníků byl využit postup uvedený v kapitole 4.3.2. Z predikcí očekávaného počtu transakcí vyplývá zvýšení přesnosti modelu při využití více dat ke kalibraci modelu (viz kapitola 5.5.2). Proto pro predikce CLV hodnot do budoucna je Pareto/NBD model kalibrován s využitím dat z celého sledovaného období (232 týdnů). Gamma-gamma model je rovněž kalibrován pro toto období.

Na Obr. 13 jsou zobrazeny predikce CLV hodnot zákazníků odhadované na následující tři roky po 232. týdnu. Zákazníci jsou seskupeni do jednotlivých segmentů podle jejich CLV hodnoty.



Obr. 13 - Predikce CLV pro následující tři roky Pareto/NBD model (Zdroj: autor)

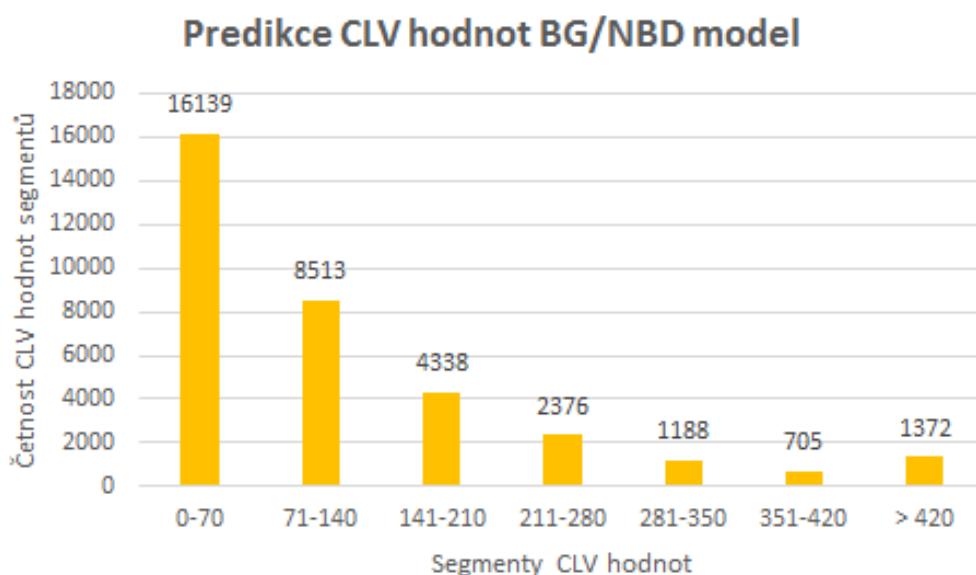
Mezi celkovým počtem 34631 zákazníků lze tedy sledovat rozdílné hodnoty CLV jednotlivých zákazníků. Charakteristika zastoupení jednotlivých hodnot je v následujících bodech.

- průměrná hodnota CLV: 359,60 Kč
- medián predikovaných CLV hodnot: 217,5 Kč
- maximální hodnota CLV: 15 700 Kč
- minimální hodnota CLV: 0 Kč

Tyto hodnoty tak představují predikce CLV hodnot za následující tři roky po 232. týdnu. Nulové hodnoty CLV připadají na zákazníky, kteří mají nízkou hodnotu frekvence realizovaných transakcí a kteří nerealizovali objednávky v nedávné době vzhledem ke konci kalibračního období (232. týden).

6.1.2. Predikce CLV pomocí BG/NBD modelu

Aplikováním stejného postupu byly provedeny predikce CLV hodnot pomocí BG/NBD modelu. Výsledky tohoto modelu zachycuje Obr. 14.



Obr. 14 - Predikce CLV pro následující tři roky BG/NBD model (Zdroj: autor)

BG/NBD model poskytuje znatelně nižší hodnoty predikcí CLV. Základní hodnoty pro tento model jsou následující.

- průměrná hodnota CLV: 123,60 Kč
- medián predikovaných CLV hodnot: 76,30 Kč
- maximální hodnota CLV: 5 927 Kč
- minimální hodnota CLV: 0 Kč

Z porovnání výsledků modelů tak vyplývá, že Pareto/NBD model poskytuje přijatelné hodnoty, zatímco BG/NBD model podhodnocuje dlouhodobou hodnotu zákazníků. Tento

model podhodnocuje očekávaný počet objednávek (viz kapitola 5.5.2) a následně proto poskytuje i nízké hodnoty CLV. Názornější příklad lze pozorovat na predikcích pro individuální zákazníky v kapitole 6.3.

6.2. Predikce CE

K vyjádření CE je použita zjednodušená definice, která určuje tuto hodnotu jako součet CLV hodnot celkové zákaznické báze (viz kapitola 3.1). Zde je nutné zmínit, že predikce proto neobsahují CLV hodnoty nově získaných zákazníků.

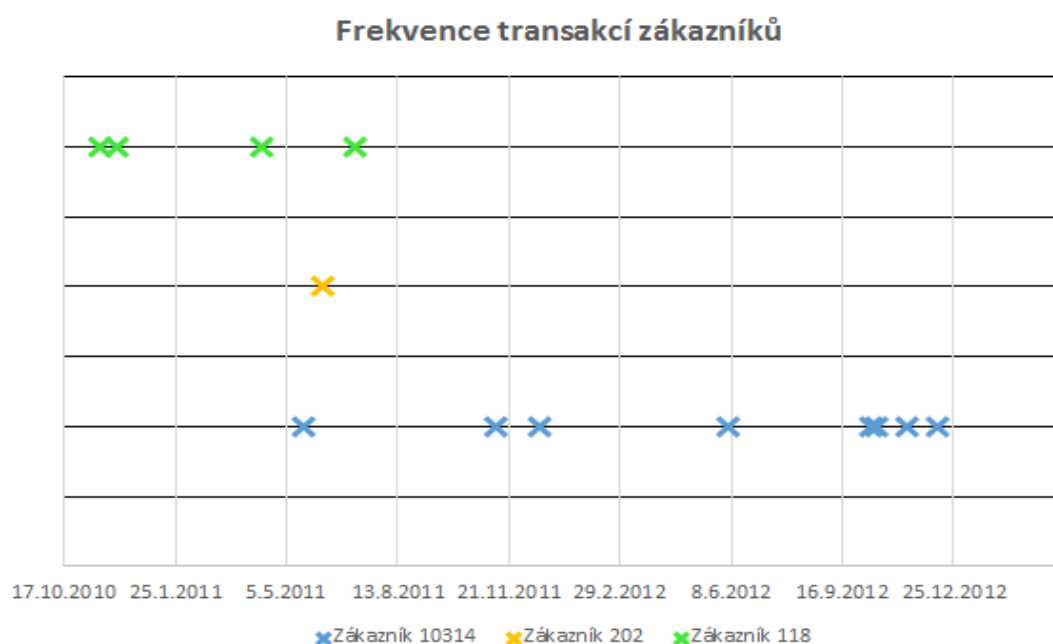
Využitím predikcí CLV hodnot zákazníků pomocí Pareto/NBD a BG/NBD modelu lze vyjádřit i hodnotu zákaznické báze pro následující tři roky po 232. týdnu. Tyto hodnoty jsou pro jednotlivé modely odlišné, protože BG/NBD model nabízí i nižší výsledky individuálních predikcí CLV.

Využitím Pareto/NBD modelu je tak predikovaná hodnota CE 12 451 748 Kč za následující tři roky.

Výsledkem BG/NBD modelu je hodnota CE 4 281 409 Kč. Nižší predikce jednotlivých CLV hodnot a následně tedy i CE pomocí BG/NBD modelu je způsobena tím, že tento model poskytuje nižší očekávaný počet transakcí a nižší hodnotu pravděpodobnosti, že bude zákazník aktivní po následující tři roky.

6.3. Srovnání predikcí modelů pro jednotlivé zákazníky

Pro srovnání predikcí CLV jednotlivých zákazníků, byli vybráni zákazníci s rozdílným chováním během kalibračního období při rozdělení 50/50. Jejich frekvence transakcí během tohoto období (1-116. týden) zachycuje Obr. 15.



Obr. 15 - Zákazníci s rozdílnou frekvencí objednávek (Zdroj: autor)

Zákazník 202 realizoval pouze jednu objednávku již během začátku kalibračního období. Pro tohoto zákazníka lze vzhledem k jeho nízké frekvenci realizovaných objednávek realizovaných již před delší dobou očekávat, že po tomto období bude realizovat nízký počet objednávek. Dále lze očekávat nízkou pravděpodobnost, že bude aktivním i na konci tohoto období.

Zákazník 118 má oproti zákazníkovi 202 větší frekvenci, ale také realizoval objednávky již před delší dobou. Pro tohoto zákazníka lze tak očekávat trochu pozitivnější výsledky.

Zákazník 10314 v porovnání s předchozími zákazníky realizoval objednávky daleko častěji a v mnohem bližší době ke konci sledovaného období. Pro tohoto zákazníka tak lze očekávat větší pravděpodobnost, že bude aktivní i vyšší počet realizovaných objednávek.

Tabulka 7 zachycuje srovnání predikcí pro tyto zákazníky. Pro jednotlivé zákazníky je možné sledovat predikce očekávaného počtu objednávek v následujících 52 týdnech, pravděpodobnost aktivity na konci 116. týdne a predikce jejich CLV hodnot.

Tabulka 7 - Srovnání predikcí pro vybrané zákazníky (Zdroj: autor)

Pareto/NBD	Zákazník 202	Zákazník 118	Zákazník 10314
Počet objednávek	0,0636	0,522	2,017
Pravděpodobnost aktivity	40,34%	58,14 %	88,49 %
CLV	23,33 Kč	565,72 Kč	17884,38 Kč
BG/NBD	Zákazník 202	Zákazník 118	Zákazník 10314
Počet objednávek	0,0815	0,1676	0,9455
Pravděpodobnost aktivity	100 %	17,4 %	45,05 %
CLV	4,91 Kč	15,74 Kč	3728,82 Kč

Zde je důležité zdůraznit predikce BG/NBD modelem. Pravděpodobnost zákazníka 202 je 100%, protože tento zákazník realizoval jen jednu objednávku. Jak bylo zmíněno v kapitole 5.5.1, paretovské modely zkoumají opakované objednávky. Zákazník 202 má nulovou hodnotu frekvence a BG/NBD model pro tyto zákazníky predikuje sto procentní pravděpodobnost, že budou aktivní. Přesto však tento model pro ostatní zákazníky značně podhodnocuje počet objednávek a poskytuje tak nízké hodnoty predikcí CLV.

Celkem lze shrnout, že predikce odpovídají očekávaným výsledkům. Zákazník 202 nejnížší hodnotu. Pro zákazníka 10314 naopak byly predikovány nejvyšší hodnoty. Na druhé straně výsledky BG/NBD modelu podhodnocují hodnotu zákazníků, proto by pro využití predikcí CLV měl být využit Pareto/NBD model.

6.4. Zhodnocení možných dopadů predikcí na manažerská rozhodování

Využití paretovských modelů nabízí odpovědi na hlavní manažerské otázky. Pomocí predikcí lze získat informace o tom, s jakou pravděpodobností lze očekávat, že budou zákazníci v určitém období aktivní. Dále je možné predikovat, kolik lze očekávat transakcí v budoucnu pro jednotlivé zákazníky. Na základě těchto informací byly predikovány hodnoty CLV a CE, podle kterých se může podnik rozhodovat o profitabilitě jednotlivých zákazníků a ocenit hodnotu zákaznické báze.

Z kapitol 5.5 a 6 vyplývá, že BG/NBD model poskytuje nízké hodnoty predikcí. Jeho využití by vedlo v tomto případě k špatnému ohodnocení zákazníků. Podnik by se tak nezaměřoval na zákazníky, kteří ve skutečnosti poskytují vyšší hodnotu. Na druhé straně Pareto/NBD model poskytuje přesně výsledky a lze ho proto využít pro manažerská rozhodování. Predice Pareto/NBD modelu lze využít následovně:

Při poklesu očekávaného počtu transakcí a pravděpodobnosti aktivity jednotlivých zákazníků by mělo vedení podniku zvážit, zda by bylo vhodné snažit se některé z těchto zákazníků získat zpět. Pomocí predikcí CLV se může podnik naopak rozhodovat o tom, jak přistupovat k jednotlivým zákazníkům. Jak bylo ukázáno v kapitole 6.1, zákazníky lze segmentovat podle jejich hodnoty CLV. Pro jednotlivé segmenty může podnik cílit své marketingové plány a kampaně, které lze lépe přizpůsobit pro daný segment zákazníků. V kapitole 3.3 bylo zmíněno využití znalosti CLV zákazníků v CRM. Pro zákazníky s určitou hodnotou je vhodné přizpůsobit i formu komunikace s podnikem. Pro zákazníky s nižší hodnotou je vhodnější využít elektronické komunikace, čímž lze dosáhnout oslovení více zákazníků za kratší čas. Zatímco pro zákazníky s vysokou hodnotou CLV by měl podnik přistupovat osobněji například formou osobní komunikace. V prostředí internetového obchodu je možné nabízet zákazníkům doplňky a příslušenství pro objednávané zboží. Takové nabídky jsou vhodné i zejména při snížení aktivity zákazníků případně i nabídky podobného zboží.

Dále lze využít predikce CE k měření hodnoty podniku. Predikce CE slouží zejména pro vrcholné řízení, kdy je potřeba určit hodnotu celé zákaznické báze podniku. Lze sledovat vývoj této hodnoty za jednotlivá období a vyhodnocovat změny v této hodnotě. Vedení podniku může sledovat, zda klesá hodnota CE a reagovat tak vytvářením nových marketingových plánů a kampaní pro získání nových zákazníků a zvýšení hodnoty zákaznické báze.

6.5. Shrnutí

Práce se věnovala predikcím CLV v non-contractual prostředí. Pro predikce byly využity Pareto/NBD a BG/NBD modely, kterými byly predikovány očekávané počty transakcí a pravděpodobnost aktivity zákazníků. Využitím těchto informací společně s průměrnými hodnotami transakcí získaných gamma-gamma modelem byly predikovány CLV hodnoty zákazníků.

Jednotlivé predikce Pareto/NBD a BG/NBD modelů byly navzájem srovnány. Kromě srovnání výsledků predikcí je vhodné poukázat i na rychlost odhadu parametrů těmito modely. Pro srovnání času potřebného k odhadu parametrů jsem využil notebook Dell Inspiron N5110 a Lenovo ThinkPad Edge E530. Srovnání jsem provedl při rozdělení datasetu 50/50. Základní specifikace a potřebný čas k odhadu parametrů jsou následující:

Dell Inspiron N5110:

Operační systém: Windows 7 Home Premium - 64bit
Procesor: Intel Core i5-2450M CPU 2.50 GHz
Operační paměť: 6 GB
Odhad parametrů BG/NBD modelu: 14,1711 vteřin
Odhad parametrů Pareto/NBD: 1,4066 hodin

Lenovo ThinkPad Edge E530:

Operační systém: Windows 7 Professional - 64bit
Procesor: Intel Core i7-3632QM CPU 2.20 GHz
Operační paměť: 12 GB
Odhad parametrů BG/NBD modelu: 12,9357 vteřin
Odhad parametrů Pareto/NBD: 1,163 hodin

Potřebný čas k odhadu parametrů dokládá, že BG/NBD model byl vyvinut pro snazší implementaci a je pro něj jednodušší odhadnout jeho parametry. Tento faktor za předpokladu, že by BG/NBD model nepodhodnocoval hodnoty zákazníků, mohl přispět ve prospěch využití tohoto modelu.

7. Závěr

Tato práce poskytla celkový pohled na predikce dlouhodobé hodnoty zákazníka s využitím paretovských modelů. Práce se zabývala predikcemi pro non-contractual prostředí internetového obchodu, pro který byly provedeny predikce CLV a CE.

7.1. Zhodnocení dosažení stanovených cílů

Přehled splnění cílů obsahuje Tabulka 8 a dále jsou podrobněji popsány.

Tabulka 8 - Naplnění cílů práce (Zdroj: autor)

Stanovený cíl	Splnění
1. Analyzovat využívané modely pro predikci CLV	2. Rešerše zdrojů 5.5. Zhodnocení modelů
2. Simulovat predikce CLV na zákaznických datech	6. Predikce CLV
3. Kritické zhodnocení predikcí k možným dopadům na manažerská rozhodování v různých úrovních řízení	6.4. Zhodnocení možných dopadů predikcí na manažerská rozhodování
4. Rešerše zdrojů věnujících se konceptu CLV a nejčastějším způsobům predikcí za daných situací	2. Rešerše zdrojů
5. Analyzovat vhodnost vybraných modelů	5.5. Zhodnocení modelů
6. Aplikace zavedené metodiky na práci s daty pro predikce CLV	5. Praktická část – aplikování metodiky CRISP-DM

Pro splnění prvního cíle byly nejprve na základě rešerše vybrány paretovské modely pro non-contractual prostředí. Jejich celkové zhodnocení bylo provedeno společně s dosažením pátého cíle v kapitole 5.5.

K dosažení druhého cíle byl využit dataset internetového obchodu, na kterém byly provedeny predikce CLV a CE pomocí paretovských modelů. Bylo tak zjištěno, že Pareto/NBD model poskytuje přesné výsledky, zatímco BG/NBD model predikuje nízké hodnoty zákazníků (viz kapitoly 6.1 a 6.2). Využití tohoto modelu by vedlo k podhodnocování CLV hodnot zákazníků a vyjádření CE pro celou zákaznickou bázi.

Další cíl byl splněn v kapitole 6.4, kde byly popsány možnosti využití Pareto/NBD modelu a nevhodnost využití BG/NBD modelu k manažerským rozhodováním.

Provedení rešerše zdrojů věnujících se konceptu CLV bylo splněno již v kapitole 2. V této kapitole byly pro účely této práce a vybrány Pareto/NBD a BG/NBD modely, které bylo možné využít pro predikce s využitím dat internetového obchodu.

Analýzování a zhodnocení Pareto/NBD a BG/NBD modelů bylo provedeno v kapitole 5.5. Splněním tohoto cíle bylo zjištěno, že Pareto/NBD model je vhodnější pro predikce CLV na datasetu internetového obchodu, protože BG/NBD model podhodnocuje hodnotu zákazníků.

Aplikování metodiky CRISP-DM bylo provedeno v kapitole 5., kde se práce zabývala jednotlivými fázemi této metodiky při práci s datasetem internetového obchodu.

Kromě stanovených cílů byl v této práci také popsán proces predikce CLV, kterému je věnována kapitola 4.3.2. Dále se práce věnovala vztahu konceptu CLV k CRM (kapitola 3.3). Byly zde představeny důležité faktory ovlivňující predikce CLV jako například datová kvalita, životní cyklus zákazníka, porozumění hodnotě zákazníka a také výhody plynoucí z využití konceptu CLV (viz kapitola 3.6). Následně tak byly splněny cíle této práce.

7.2. Přínosy práce

Hlavními přínosy práce jsou zejména provedené predikce zákaznického chování a využití těchto informací pro odhad CLV. Práce tak poskytuje srovnání predikcí pro Pareto/NBD a BG/NBD modely a jejich výsledky lze srovnávat navzájem. Bylo zjištěno, že pro využití reálných dat z internetového obchodu je vhodný Pareto/NBD model. Dalším přínosem práce je aplikování metodiky CRISP-DM. Práce poskytuje praktický příklad využití této metodiky a věnuje se jednotlivým fázím metodiky s využitím reálných dat. Na základě predikcí Pareto/NBD modelem byla tak odhadnuta průměrná hodnota CLV 360 Kč a hodnota CE 12 451 748 Kč.

7.3. Možnosti využití a dalšího rozšíření práce

Práce poskytla srovnání Pareto/NBD a BG/NBD modelů pro predikce CLV. Realizované predikce byly zjednodušeny zejména o stanovení výše započítávaných nákladů. Možností rozšíření této práce by tedy mohlo být realizování těchto predikcí s přesnějším započítáním nákladů. Dalším případným využitím a rozšířením práce by mohlo být srovnání modelů na jiných reálných datech pro non-contractual prostředí, kde by například mohlo být vyšší zastoupení zákazníků s opakovanými transakcemi.

Terminologický slovník

Termín	Zkratka	Význam
Customer lifetime value	CLV	Současná hodnota budoucích net cash flows spojených s konkrétním zákazníkem Fader (2012)
Customer equity	CE	Suma dlouhodobých hodnot zákazníků napříč celou zákaznickou bází podniku. Fader (2012)
Customer relationship management	CRM	Řízení vztahu se zákazníky (Gála et al., 2009)
Non-contractual prostředí		Business prostředí, kde zákazník není vázán k užívání produktů či služeb žádným smluvním vztahem. Fader (2012)
Pareto / Negative Binomial Distribution model	Pareto/NBD	Pravděpodobnostní model pro predikci zákaznického chování vyvinut Schmittlein et al.(1987).
Beta-Geometric / Negative Distribution model	BG/NBD	Pravděpodobnostní model pro predikci zákaznického chování, který byl vyvinut jako alternativa k Pareto/NBD. Fader (2005a)
Gamma-gamma model	Gamma-gamma	Model pro vyjádření peněžní hodnoty transakcí. Fader (2005a)
Calibration period		Kalibrační období, které zahrnuje část určitou část dat pro kalibraci modelu. (autor)
Holdout period		Testovací (validační) období, které zahrnuje zbylou část dat pro otestování modelu. (autor)

Seznam literatury

[BAGGOT, 2013]

BAGGOTT, Matthew. *Predicting Customer Behavior with R: Part 1* [online]. 2013[cit. 2015-04-04]. Dostupné z: <http://www.slideshare.net/mattbagg/baggott-predict-customerinrpart1>

[BERGER, 1998]

BERGER, Paul D. a Nada I. NASR. Customer lifetime value: Marketing models and applications. [online]. 1998, s. 161-180 [cit. 2015-02-28]. DOI: 10.1007/978-0-387-72579-6_7. Dostupné z: http://download.clib.psu.ac.th/datawebclib/e_resource/trial_database/WileyInterScienceCD/pdf/DIR/DIR_2.pdf

[CLARK, 2010]

CLARK, Peter. The 16 business benefits of Customer Lifetime Value. [online]. The Wise Marketer, 2010 [cit. 2015-03-13]. Dostupné z: <http://www.thewisemarketer.com/features/read.asp?id=119>

[DIENER, 2013]

DIENER, Chris. Real-Time CRM To Maximize Your Customer Lifetime Value. *CMO* [online]. 2013 [cit. 2015-03-09]. Dostupné z: http://www.cmo.com/articles/2013/8/1/real_time_crm_to_max.html

[DZIURZYNSKI, 2014]

DZIURZYNSKI, L., WADSWORTH, E. a D. MCCARTHY (2014). BTYD: Implementing Buy 'Til You Die Models. R package version 2.4. <http://CRAN.R-project.org/package=BTYD>

[FADER, 2005a]

FADER, Peter S., Bruce G. S. HARDIE a K. L. LEE. RFM and CLV: Using Iso-value Curves for Customer Base Analysis. *American Marketing Association* [online]. 2005, s. 415-430 [cit. 2015-04-07]. Dostupné z: http://brucehardie.com/papers/rfm_clv_2005-02-16.pdf

[FADER, 2012]

FADER, Peter. *Customer centricity: Focus on the Right Customers for Strategic Advantage*. 2. vyd. Philadelphia: Wharton Digital Press, 2012. ISBN 978-1-61363-015-0.

[FADER, 2009]

FADER, Peter S. a Bruce G.S. HARDIE. Probability Models for Customer-Base Analysis. [online]. 2009, s. 61-69 [cit. 2015-03-03]. DOI: 10.1016/j.intmar.2008.11.003. Dostupné z: <https://marketing.wharton.upenn.edu/index.cfm/research/research-listing/?whdmsaction=publications.list&pubFilter=publishedPaper&pubYearFilter=2009>

[Fader, 2008]

FADER, Peter S., Bruce G. S. HARDIE a Ka L. LEE. Computing P(alive) Using the BG/NBD Model. In: [online]. 2008 [cit. 2015-04-27]. Dostupné z: <http://brucehardie.com/notes/>

[Fader, 2005b]

FADER, Peter S., Bruce G. S. HARDIE a Ka Lok LEE. "Counting Your Customers" the Easy Way: An Alternative to the Pareto/NBD Model. *Marketing Science* [online]. 2005, s. 275-284 [cit. 2015-04-01]. DOI: 10.1287. Dostupné z: https://www.google.cz/url?sa=t&rct=j&q=&esrc=s&source=web&cd=3&ved=0CDgQFjAC&url=https%3A%2F%2Fmarketing.wharton.upenn.edu%2Ffiles%2F%3Fwhdmsaction%3Dpublic%3Amain.file%26fileID%3D4240&ei=2ggjVZb_NNHVaqm3glAE&usg=AFQjCNFhjyJC48GZBV1zOr2EHG92Bs7rTQ&sig2=z1_Cr7Ss-6ULogOb7nwFKA&bvm=bv.89947451,d.d2s&cad=rja

[GÁLA et al., 2009]

GÁLA, Libor, Jan POUR a Zuzana ŠEDIVÁ. *Podniková informatika*. 2., přeprac. a aktualiz. vyd. Praha: Grada, 2009, 496 s. Expert (Grada). ISBN 9788024726151.

[GUPTA, 2005]

GUPTA, S a D. R. LEHMANN. *Managing customers as investments: the strategic value of customers in the long run*. Upper Saddle River, NJ: Wharton School Pub., 2005, 205 s. ISBN 0-13-142895-0.

[GUPTA et al., 2006]

GUPTA, S. et al Modeling Customer Lifetime Value. [online]. 2006, s. 139-155 [cit. 2015-02-27]. DOI: 10.1177/1094670506293810. Dostupné z: <http://www.anderson.ucla.edu/faculty/dominique.hanssens/content/JSR2006.pdf>

[CHAPMAN et al., 2000]

CHAPMAN, P., J. CLINTON, R. KERBER, T. KHABAZA, C. SHEARER a R. WIRTH. 2000. *CRISP-DM 1.0* [online]. [cit. 2015-04-01]. Dostupné z: <http://the-modeling-agency.com/crisp-dm.pdf>

[JAIN, 2002]

JAIN, Dipak a Siddhartha S. SINGH. Customer lifetime value research in marketing: A review and future directions. [online]. Direct Marketing Educational Foundation, Inc., 2002, s. 118-136 [cit. 2015-03-13]. DOI: 10.1002/dir.10032. Dostupné z: http://www.researchgate.net/profile/Siddharth_Singh8/publication/230537924_Customer_lifetime_value_research_in_marketing_A_review_and_future_directions/links/0046351a7f83555b2d000000.pdf

[MACHANDER, 2012]

MACHANDER, Aleš. *Aplikace CVM v telekomunikační společnosti*. Praha, 2012. Diplomová práce. Vysoká škola ekonomická. Vedoucí práce doc. Ing. Ota Novotný, Ph.D.

[MCCARTHY, 2002]

MCCARTHY, Daniel a Edward WADSWORTH. *Buy 'Til You Die - A Walkthrough* [online]. 2014 [cit. 2015-04-04]. Dostupné z: <http://cran.r-project.org/web/packages/BTYD/vignettes/BTYD-walkthrough.pdf>

[NOVO, 2004]

NOVO, Jim. *Drilling Down: Turning Customer Data Into Profits with a Spreadsheet*. 3. vyd. Saint Petersburg: Booklocker.com, Inc., 2004. ISBN 1-59113-519-2.

[OLSON, 2003]

OLSON, Jack E. *Data Quality: The Accuracy Dimension*. San Francisco: Morgan Kaufman Publishers, 2003. ISBN 1-55860-891-5.

[PAVEL, 2012]

PAVEL, Jan. *Customer Intelligence v prostředí elektronického obchodu*. Praha, 2012. Diplomová práce. Vysoká škola ekonomická. Vedoucí práce Ing. Ota Novotný, Ph.D.

[PLATZER, 2008]

PLATZER, Michael. *Stochastic Models of Noncontractual Consumer Relationships*. Vídeň, 2008. Diplomová práce. Vienna University of Economics and Business Administration. Vedoucí práce Dr. Thomas Reutterer.

[R Core Team, 2015]

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

[ROŽEK, 2014]

ROŽEK, J. a M. KARLÍČEK. Customer Lifetime Value as the 21st century Marketing Strategy Approach. *Central European Business Review* [online]. 2014, s. 28-35 [cit. 2015-03-14]. Dostupné z: <http://oaji.net/articles/2014/1215-1409577805.pdf>

[SCHMITTLEIN et al., 1987]

SCHMITTLEIN, David C., Donald G. MORRISON a Richard COLOMBO. Counting your customers: who are they and what will they do next?. *Management Science*. 1987, 1 - 24.

[WÜBBEN, 2008]

M., Wübben a Von Wangenheim F. Instant Customer Base Analysis: Managerial Heuristics Often "Get It Right". *Journal of Marketing* [online]. 2008, s. 82-93 [cit. 2015-04-28]. Dostupné z: http://wuebben.com/members/markus/Home/Publications_files/wuebben%20wangenheim%20JM%202008%20instant%20customer%20base%20analysis.pdf

Seznam obrázků a tabulek

Obr. 1 - Transakční historie (Fader, 2009).....	11
Obr. 2 - Klasifikace vztahů se zákazníky (Fader, 2009).....	11
Obr. 3 - Architektura CRM - Gála et. al (2009)	14
Obr. 4 - Životní cyklus zákazníka (Rožek, Karlíček 2014).....	16
Obr. 5 - Hodnota zákaznického portfolia (Novo, 2004)	17
Obr. 6 - Hodnota zákazníka pro podnik a hodnota pro zákazníka (Gupta, 2005)	19
Obr. 7 - Pravděpodobnost aktivity zákazníků na konci kalibračního období (Zdroj: autor)	31
Obr. 8 - Pareto/NBD s rozdělením 50/50 (Zdroj: autor)	32
Obr. 9 - Pareto/NBD s rozdělením 90/10 (Zdroj: autor)	33
Obr. 10 - BG/NBD s rozdělením 50/50 (Zdroj: autor).....	34
Obr. 11 - BG/NBD s rozdělením 90/10 (Zdroj: autor).....	34
Obr. 12 - Gamma-gamma model zhodnocení predikcí (Zdroj: autor)	36
Obr. 13 - Predikce CLV pro následující tři roky Pareto/NBD model (Zdroj: autor)	37
Obr. 14 - Predikce CLV pro následující tři roky BG/NBD model (Zdroj: autor).....	38
Obr. 15 - Zákazníci s rozdílnou frekvencí objednávek (Zdroj: autor)	39
Tabulka 1 - Množství objednávek pro vybrané zákazníky (Zdroj: autor)	26
Tabulka 2 - Charakteristika datasetu (Zdroj: autor)	26
Tabulka 3 - Odhad parametrů (Zdroj: autor - realizováno v R).....	30
Tabulka 4 - Predikce počtu transakcí Pareto/NBD modelem (Zdroj: autor - realizováno v R)	33
Tabulka 5 - Predikce počtu transakcí BG/NBD modelem (Zdroj: autor - realizováno v R)	35
Tabulka 6 - Zhodnocení modelů RMSE a MSLE (Zdroj: autor - realizováno v R).....	35
Tabulka 7 - Srovnání predikcí pro vybrané zákazníky (Zdroj: autor).....	40
Tabulka 8 - Naplnění cílů práce (Zdroj: autor)	43