

JIHOČESKÁ UNIVERZITA V ČESKÝCH BUDĚJOVICÍCH

EKONOMICKÁ FAKULTA

KATEDRA APLIKOVANÉ MATEMATIKY A
INFORMATIKY

STUDIJNÍ PROGRAM: Ekonomika a management

STUDIJNÍ OBOR: Strukturální politika EU a rozvoj venkova



VYUŽITÍ KLASIFIKAČNÍCH STROMŮ V MARKETINGOVÉM PRŮZKUMU TRHU

Vedoucí diplomové práce

Ing. Michael Rost, Ph.D.

Autor

Bc. Kateřina Prokopová

2008

JIHOČESKÁ UNIVERZITA V ČESKÝCH BUDĚJOVICÍCH

Ekonomická fakulta

Katedra aplikované matematiky a informatiky

Akademický rok: 2006/2007

ZADÁNÍ DIPLOMOVÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Bc. Kateřina PROKOPOVÁ**

Studijní program: **N6208 Ekonomika a management**

Studijní obor: **Strukturální politika EU a rozvoj venkova**

Název tématu: **Využití klasifikačních stromů v marketingovém průzkumu trhu**

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

Cíl práce:

Pomocí progresivní metodologie CART (klasifikační a regresní stromy) identifikovat důležité faktory ovlivňující chování zákazníka.

Metodický postup:

Studiem literatury budou získány znalosti o principu klasifikačních metod. Zvláštní pozornost bude věnována metodologii CART. Na zvoleném příkladě marketingového průzkumu trhu budou prostřednictvím klasifikačních stromů identifikovány faktory ovlivňující spotřební chování. K samotnému výpočtu bude použito programovací prostředí R případně STATISTICA.

Hlavní body metodického postupu:

1. Studium literatury.
2. Sběr dat případně vlastní primární výzkum.
3. Statistické zhodnocení získaných dat prostřednictvím metodologie CART.
4. Formulace závěrů.

Rozsah grafických prací: dle potřeby
Rozsah pracovní zprávy: 50 - 70 stran
Forma zpracování diplomové práce: tištěná

Seznam odborné literatury:

BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R. A, STONE, CH. J.
Classification and Regression trees, Chapman & Hall/CRC, Boca Raton,
1998, 358 p., ISBN 0-412-04841-8.

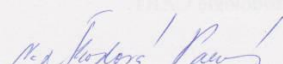
HEBÁK, P., HUSTOPECKÝ, J. Vícerozměrné statistické metody s apli-
kacemi, SNTL, Praha, 1987, 438 s.

R Development Core Team (2004). R: A language and environment for
statistical computing. R Foundation for Statistical Computing, Vienna,
Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.

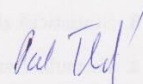
Vedoucí diplomové práce: Ing. Michael Rost, Ph.D.
Katedra aplikované matematiky a informatiky

Datum zadání diplomové práce: 10. ledna 2007

Termín odevzdání diplomové práce: 15. dubna 2008


prof. Ing. Magdalena Hrabánková, CSc.
děkanka

JIHOČESKÁ UNIVERZITA
V ČESKÝCH BUDĚJOVICÍCH
EKONOMICKÁ FAKULTA
S. J. 13 (1)
370 05 České Budějovice


doc. RNDr. Pavel Tlustý, CSc.
vedoucí katedry

V Českých Budějovicích dne 12. března 2007

Prohlášení

Prohlašuji, že jsem diplomovou práci na téma Využití klasifikačních stromů v marketingovém průzkumu trhu vypracovala samostatně na základě vlastních zjištění a materiálů, které uvádím v seznamu použité literatury.

V Třebíči 25. 4. 2008

Bc. Kateřina Prokopová

Poděkování

Děkuji panu Ing. Michaelovi Rostovi, Ph.D., vedoucímu mé diplomové práce, za pomoc a rady, které mi při zpracování této diplomové práce poskytl.

OBSAH:

1. Úvod	4
2. Přehled literatury	5
2.1 Mobilní operátoři v České republice	5
2.1.1 Vodafone Czech republic, a.s.	5
2.1.2 T-Mobile Czech Republic, a.s.	7
2.1.3 Telefónica O2 Czech Republic, a.s.	9
2.2 Rozhodovací stromy	11
2.2.1 Historie	11
2.2.2 Klasifikační stromy	12
3. Metodika	19
4. Vlastní zpracování	21
4.1 Vyhodnocení dotazníků	21
4.2 Sestrojení klasifikačních stromů	33
4.3 Diskuse	45
5. Závěr	58
6. Summary	60
7. Použitá literatura	61
8. Seznam obrázků, tabulek a grafů	62
8. Přílohy	65

1. ÚVOD:

Ve své diplomové práci se budu zabývat využitím klasifikačních stromů v marketingovém průzkumu trhu, přičemž se zaměřím na oblast poskytování mobilních služeb. V dnešní době je to velmi aktuální téma, protože kdo nemá mobilní telefon, jako by nebyl. Právě tohoto moderního celospolečenského trendu chtějí co nejvíce využít všichni mobilní operátoři. Každá ze společností se snaží co nejvíce zvýšit počet svých uživatelů a zároveň udělat vše proto, aby stálí zákazníci neměli důvod odcházet jinam.

Cílem mé práce je pomocí progresivní metodologie CART (klasifikační a regresní stromy) identifikovat důležité faktory ovlivňující chování zákazníka.

V rámci metodiky budu postupovat od teoretických poznatků ke statistickému zhodnocení získaných dat prostřednictvím metodologie CART.

V teoretické části diplomové práce se zaměřím na základní informace týkajících se tří největších mobilních operátorů u nás, a také na problematiku klasifikačních stromů. Z literatury, ze které budu čerpat, bych chtěla především uvést Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C. J.: Classification and Regression Trees.

V praktické části diplomové práce budu nejdříve provádět sběr informací týkajících se mobilních operátorů v ČR na základě dotazníkového šetření. Na základě získaných dat budu pomocí klasifikačních stromů identifikovat faktory, kterými jsou zákazníci při výběru mobilního operátora nejvíce ovlivněni. K vyhodnocení dotazníků použiji programovací prostředí R a Excel.

2. PŘEHLED LITERATURY:

2.1 MOBILNÍ OPERÁTOŘI V ČESKÉ REPUBLICE:

V České republice existují čtyři mobilní operátoři: Telefónica O2 Czech Republic, a.s., T-Mobile Czech Republic, a.s., Vodafone Czech Republic, a.s., a nově mobilní operátor U:fon – MobilKom, a.s.. Ve své diplomové práci se ovšem zaměřím pouze na tři největší mobilní operátory u nás.



2.1.1 Vodafone Czech republic, a.s.:

Po získání licence na provozování duální mobilní sítě GSM 900/1800 v říjnu roku 1990 se Český Mobil, a.s. stal prvním duálním mobilním operátorem v České republice. Komerční provoz zahájil pod značkou Oskar. 1. března 2000 vstoupil na trh jako třetí a nejrychleji se rozvíjející mobilní operátor v České republice. Zasadil se o radikální změnu na českém mobilním komunikačním trhu a mobilní telefony cenově zpřístupnil každému spotřebiteli. V červnu roku 2005 se stal členem rodiny Vodafone. Vodafone je největší mobilní operátor na světě. Působí v 28 zemích pěti kontinentů [1].

Základní fakta a čísla:

- Počet zákazníků: 2.263 miliónů (k 30.6. 2006)
- Hlavní výhody služeb sítě Vodafone:
 - Žádný aktivační poplatek
 - Žádný smluvní závazek

- Jedna cena do všech sítí po celý den
 - SMS za 1.19 Kč (včetně DPH) do všech sítí
 - Účtování hovorného po vteřinách po první minutě
 - Převod neprovolaných volných minut do následujícího měsíce
- Pokrytí: Síť Vodafone pokrývá 99% populace (k 31.12.2005)

Poskytované služby:

- a) Vodafone nabízí 5 *nabitých tarifů* v cenovém rozpětí od 119Kč do 2100Kč. Zákazník si může zvolit, zda měsíční paušál využije na volné minuty nebo SMS zprávy. Např. u tarifu Nabito 119 může zákazník volit mezi 19-ti volnými minutami nebo může zvolit až 100 volných SMS zpráv. Tarif Nabito 2100 nabízí buď 700 volných minut nebo až 1764 SMS zpráv.
- b) Vodafone také nabízí jeden, zákazníky velmi oblíbený tarif, a to *tarif Odepiš*. Zákazník zaplatí měsíčně pouhých 12Kč a má k dispozici 40 volných minut do sítě Vodafone. Tento tarif může být sjednán pouze jednou na jedno rodné číslo.
- c) *Vodafone divoká karta* – její výhodou je, že se neplatí žádný měsíční paušál. Zákazník si ji může dobíjet podle potřeby, a to buď nákupy dobíjecích kartiček na pokladnách, nebo prostřednictvím terminálu sazky či pomocí bankomatu.
- d) *Hlasové služby* – hlasová schránka, apod.
- e) *Roaming a volání do zahraničí*
- f) *Přijímání a posílání faxů přes mobil*
- g) *MMS služby*

- h) *SIM plus* – možnost ovládání bankovního účtu přímo z mobilu.
- i) *Internet, WAP a Vodafone life* – díky této službě má klient neustálý přístup k internetu, svým e-mailovým schránkám, Vodafone live i WAPu. K internetu tak může připojit kdykoli a kdekoli svůj mobil, notebook nebo PDA.



2.1.2 T-Mobile Czech Republic, a.s.:

T-Mobile Česká republika, a.s. je součástí mezinárodní skupiny T-mobile International, jež nabízí služby 103 miliónům zákazníků v 8 evropských zemích – v ČR, Chorvatsku, Maďarsku, Nizozemí, Německu, Rakousku, na Slovensku, ve Velké Británii a v USA. Kromě toho T-Mobile International vlastní 49% podíl ve společnosti Polska Telefonica Cryfrowa (PTC), jež je největším a jediným plně privatizovaným polským mobilním operátorem, a dále 41% podíl v Mobile TeleSystems, což je největší mobilní operátor v Rusku [2].

Historie společnosti:

Vítězem nabídkového řízení ministerstva hospodářství na mezinárodního partnera Českým radiokomunikacím se stalo dne 14. března 1996 konsorcium CMobil. Jedenáct dnů poté byla slavnostně předána Pověření k provozování sítě mobilních telefonů GSM. Přibližně za čtvrt roku byla u obchodního soudu registrována nová společnost RadioMobil, která zahájila provoz sítě mobilních telefonů Paegas 30. září 1996. V roce 2002 společnost přejmenovala síť na T-Mobile a o rok později přijala T-Mobile i jako obchodní název.

Základní fakta a čísla:

- Počet zákazníků T-Mobile Czech Republic, a.s.: 5,140 miliónu (k 30.6.2007)

- Počet zákazníků mezinárodní skupiny T-Mobile: 112 miliónů (k 30.6.2007)
- Počet roamingových partnerů: téměř 400 operátorů ve více než 160 zemích
- Pokrytí v procentech: přes 99% české populace a 99,98% silnic třídy E na území ČR

Poskytované služby:

- a) *Kreditní tarify* – v rámci měsíčního paušálu získá zákazník kredit na volání, SMS a MMS. Kredit čerpá libovolně, jak mu to právě vyhovuje a nevyužitý kredit se vždy převádí do dalšího měsíce. Zákazník může volit mezi pěti takovými tarify, a to v cenovém rozmezí 297,5Kč až 2380Kč, přičemž čím vyšší je kredit, tím je minuta hovoru levnější. Např. u nejnižšího tarifu je cena minuty 5,95Kč a u nejvyššího tarifu je cena minuty 2,98 Kč.
- b) *Minutové tarify* - Zákazník má možnost vybírat z šesti minutových tarifů, přičemž nejnižší tarif pořídí za 226,1Kč a získá 30 volných minut, nejdražší potom za 4462,5Kč s 1500 volnými minutami.



2.1.3 Telefonica O2 Czech Republic, a.s.:

Společnost Telefonica působí ve 23 zemích světa a počet jejich zákazníků dosahuje 196 miliónů. Telefonica O2 Czech Republic, a.s., je prvním integrovaným operátorem v České republice, který vznikl 1. července 2006 spojením nejvýznamnějšího provozovatele pevných linek, ČESKÉHO TELECOMU, a.s., a nejsilnějšího mobilního operátora, Eurotel Praha, spol. s.r.o., do jedné telekomunikační společnosti. Společnost dnes provozuje téměř osm miliónů mobilních a pevných linek, což z ní činí jednoho z vedoucích poskytovatelů plně konvergentních služeb na světě. Společnost nabízí nejucelenější nabídku hlasových a

datových služeb v České republice. Mimořádnou pozornost věnuje využití růstového potenciálu především v datové a internetové oblasti. Společnost provozuje nejrozsáhlejší pevnou a mobilní síť včetně unikátních sítí 3. generace – datovou síť CDMA a síť UMTS, která umožňuje přenos dat, hlasu, obrazu a videa. Telefónica O2 Czech Republic nabízí také největší síť WiFi hotspotů v zemi [3].

Poskytované služby:

- a) *Simple tarify* - Tyto tarify můžeme pořídit od 285,6Kč do 2356,2Kč a čím dražší tarif, tím je minuta hovoru levnější.
- b) *Standardní tarify* - V případě, že zákazník preferuje zejména posílání SMS zpráv může si zvolit tarif, který poskytuje 150 volných SMS zpráv za 214,2Kč. V opačném případě si může vybrat z dalších pěti tarifů, kde získá 30 až 1000 minut za 214,2Kč až 3689,2Kč.
- c) *Předplacené tarify:*
 - O2 Fajn – volání 5,50 Kč/min celý den do všech sítí, SMS za 3,20 Kč.
 - O2 TXT – předplacený tarif pro mladé. S tarifem může zákazník posílat ještě levnější SMS a MMS do všech sítí a navíc bude celý den levně volat v rámci sítě O2.
 - O2 Mix – předplacený tarif pro mladé. S tímto tarifem si zákazník namíchá poměr volání a SMS podle vlastní chuti.
- d) O2 SMSender – aplikace, díky které lze vytvářet a posílat SMS a MMS zprávy ze svého počítače na mobil.
- e) Volání po 3. minutě zdarma – u předplacených karet, do vlastní sítě.
- f) Volání domů zdarma – z mobilu na pevnou.
- g) O2 Nejbližší – umožňuje volání s blízkými výhodně a bez omezení.

- h) O2 Navigace
- i) O2 Roaming
- j) O2 Internet – společnost nabízí 5 datových tarifů v cenovém rozmezí 475 Kč do 1070 Kč podle rychlosti připojení.
- k) O2 Internet ADSL
- l) O2 Hot Spot – vysokorychlostní bezdrátové připojení k internetu.
- m) O2 Tanedem – volání si navzájem ve skupině levněji.
- n) O2 NetCall *55 – volání do ciziny výhodněji.

2.2 Rozhodovací stromy:

Metody rozhodovacích stromů mohou být neznámé statistikům, ovšem vědci z jiných oblastí používají rozhodovací stromy běžně, k vyjádření jejich znalostí. Často se používají například v botanice nebo v medicíně. Stromy umožňují zestručnit a uspořádat znalosti expertů, tak aby je mohli využívat i méně zkušení uživatelé.

2.2.1 Historie rozhodovacích stromů:

DIVIDE ET IMPERA!

- Rozděl a panuj: vhodně rozdělím zkoumané objekty do skupin...
- a v každé skupině opět postupuji stejně (rekurze)...
- dokud nedojdu k malým skupinkám, na něž stačí zcela jednoduchý model.

DIVIDE ET IMPERA je staré římské přísloví. Jeho použití v analýze dat ve smyslu rozhodovacích stromů, bylo navrženo až do roku 1959 W.A. Belson [4] (britský sociolog a metodolog, který se zabýval především kriminalitou mládeže).

Již dříve (30. léta 20. stol.) se však statistici zabývali problémy kategorizace spojitých proměnných a dělením populací, ovšem v jiném kontextu.

První počítačově implementovaný algoritmus se jmenoval AID (Morgan and Sonquist, 1963 [5]). AID byl založen na analýze rozptylu (sumy čtverců) – pomůcka pro přípravu ANOVA => základ statistického směru teorie rozhodovacích stromů (CHAID, SEARCH aj.).

Označení „rozhodovací strom“ (Decision Tree) je snad z r. 1966 (Experiments in Induction – E. B. Hunt, J. Marinová a P. J. Stone [6]) => směr zakotvený v teorii umělé inteligence. Zde vyšli z teorie informace – rozdělení na podskupiny má přinést „informační zisk“, snížit entropii.

Rozvoj aplikací a uplatnění i mimo oblast teoretické vědy přinesl nástup rychlých PC a rozvoj data miningu (cca polovina 90. let 20. století).

Mezi metody rozhodovacích stromů patří klasifikační stromy a regresní stromy. Struktura regresních stromů je jednodušší než struktura klasifikačních stromů. Regresní stromy využívají již výše zmíněný program AID (Automatic Interaction Detection). Hlavní rozdíl mezi AID a CART můžeme spatřovat v prořezávání a v procesu odhadování.

2.2.2 Klasifikační stromy:

Cílem klasifikačního stromu je sekvenčně rozdělit data tak, aby se maximalizovaly rozdíly v závislé proměnné. Tomuto postupu se často říká rozhodovací strom. Skutečným účelem klasifikačního stromu je roztrždit data do odlišných skupin či větví, které vytvoří nejsilnější separaci hodnot závislé proměnné. Klasifikační stromy jsou velmi dobré pro identifikaci segmentů s požadovaným chováním, například pro modelování odezvy či aktivace. Tato identifikace může být velice prospěšná, také když se firma snaží porozumět, co ovlivňuje chování trhu.

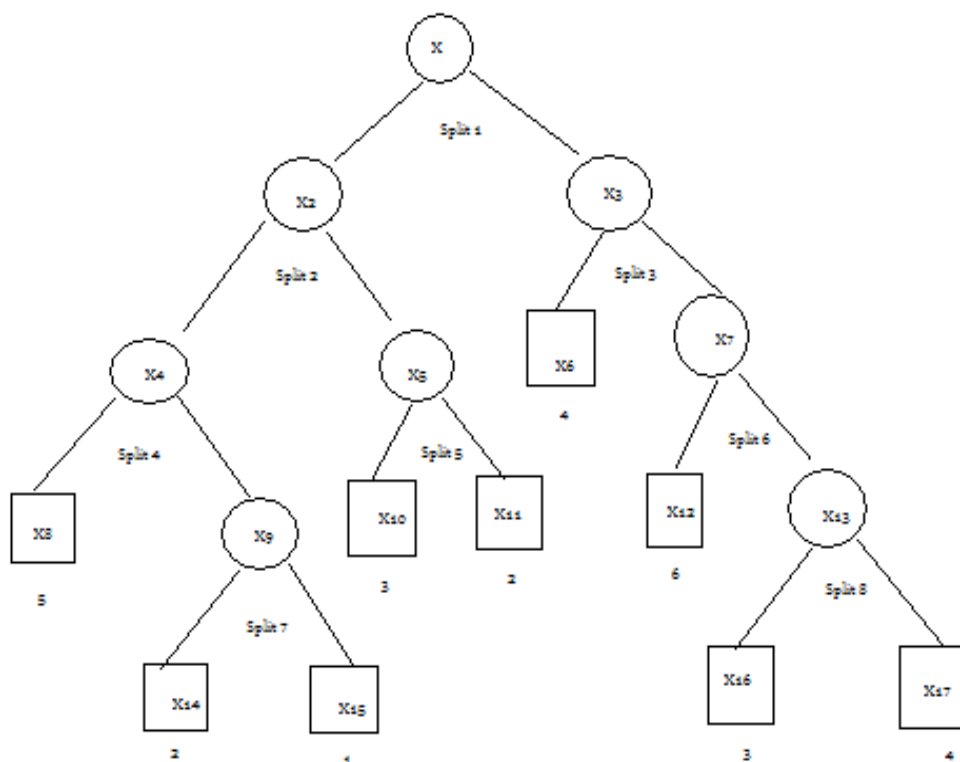
Klasifikační stromy jsou konstruovány opakovaným dělením podmnožiny X do dalších dvou podmnožin, které začínají v samotném X . Podmnožiny, které již nejsou dále rozvětvené, se nazývají konečné podmnožiny a znázorňují se pomocí pravoúhlého čtverce. Nekonečné podmnožiny se znázorňují pomocí kruhu. Každá konečná podmnožina je určena třídním označením. Můžeme mít dvě nebo více konečných podmnožin se stejným třídním označením.

Celá konstrukce stromů se točí kolem tří elementů:

- Výběr rozdělení
- Rozhodnutí o konečném uzlu
- Určení dalších konečných uzlů [7].

Na obrázku č. 1 můžeme sledovat, že pro definování prvního rozvětvení klasifikační strom určuje, zda X půjde do X_2 nebo do X_3 . Tak pokračuje dále, až se na konci X dostane do konečné podmnožiny. Tento bod můžeme nazvat konečný uzel a nekončící podmnožinu, potom nekončící uzel.

Obrázek 1: Klasifikační strom



Zdroj: Breiman a kol.

Cílem konstrukce stromu v praktických situacích bude zpravidla sestavit jeden klasifikační model, nikoli celou sérii. Obvykle však chybí nějaké jasné vodítko, podle kterého bychom předem dovedli určit správnou velikost stromu. Příliš malé stromy nepostihnou všechny zákonitosti „za daty“, příliš velké stromy budou vedle těchto zákonitostí odrážet i nahodilé vlastnosti dat.

Existují různé metody sestavení stromů. Některé přidávají ke stromu nová větvení, dokud to přináší dostatečný pokles chyby, potom je konstrukce ukončena.

Jiné metody postupují odlišně: v první fázi se také postupně přidávají nová větvení, ale kritérium, podle kterého se tento proces zastaví, je velmi mírné, s velkou pravděpodobností se vytvoří velmi velký strom. Ve druhé fázi se tento strom prořezává. Studují se stromy, které vzniknou odstraněním většího či menšího počtu větvení z původního velkého stromu (prořezané podstromy). U každého z těchto prořezaných podstromů se odhaduje kvalita generalizace a na základě těchto odhadů se pak vybere nejvhodnější velikost stromu.

Klasické metody jako CART (Breiman et al., 1984) se v zásadě snaží také konstruovat stromy s co nejmenší chybou (resp. ztrátou) na trénovacím souboru. Začnou s jediným uzlem (který je současně kořenem i listem) a postupně přidávají další uzly tak, aby se v každém kroku chování stromu na trénovacích datech v jistém smyslu co nejvíce zlepšilo. Tyto „lokálně optimální“ kroky však nezaručují, že výsledný strom bude na trénovacích datech optimální [8].

Je známo, že příliš malé stromy mají obvykle jak velkou chybu na trénovacích datech, tak i velkou chybu skutečnou, zatímco stromy příliš velké mají typicky natrénovacích datech velmi malou chybu, ale jejich skutečná chyba je o mnoho větší.

Rozhodovací stromy jsou mnohoúrovňovým systémem [9]:

- třídy jsou postupně zamítány, dokud nedosáhneme přijatelné třídy
- na konci dosáhneme stavu, kdy příznakový prostor je sekvenčně rozdělen do oblastí, které odpovídají jednotlivým třídám
- pro předložený „neznámý“ příznakový vektor, hledáme oblast, kam bude vektor zařazen pomocí posloupnosti rozhodnutí ve stromové struktuře

Nejpopulárnější jsou obecné binární rozhodovací stromy (OBCT = ordinary binary classification trees), které:

- rozdělují prostor do mnohoúhelníků, které mají strany rovnoběžné s osami

- posloupnost rozhodnutí je aplikována na jednotlivé příznaky
- rozhodování probíhá podle dotazů ve tvaru „je příznak $x_i \leq \alpha$?“

Vlastnosti binární klasifikačního stromu:

Během trénovací fáze (výstavba stromu) musíme brát v úvahu, že:

- každý uzel t má svoji trénovací množinu X_t ($X_t \subseteq X$)
- v každém uzlu máme k dispozici množinu dotazů
- každý dotaz odpovídá dělení uzlu do 2 synů
 - dělení uzlu je ekvivalentní dělení X_t do dvou disjunktních podmnožin X_{tY} a X_{tN}
 - v X_{tY} jsou vektory z X_t , na které jsme odpověděli „Yes“ pomocí dotazu v uzlu t
 - v X_{tN} jsou vektory z X_t , na které jsme odpověděli „No“ pomocí dotazu v uzlu t
 - pro každé dělení platí:

$$X_{tY} \cap X_{tN} = \emptyset$$

$$X_{tY} \cup X_{tN} = X_t$$

- dělicí kritérium musí být takové, aby vybralo nejlepší dělení z množiny kandidátů
- ukončovací podmínka definuje, kdy bude uzel listem (a tedy nedojde k dělení uzlu)
 - určuje výšku stromu
 - ke každému listu je přiřazena jeho třída

Množina dotazů :

Pro Binární klasifikační stromy jsou dotazy ve tvaru „je příznak $x_i \leq \alpha$?“. V praxi uvažujeme jen konečně dotazů:

- počet trénovacích vzorů v X je konečný (označme ho N)
- libovolný příznak x_k ($k=1, \dots, l$) může nabývat nejvýše $N_t \leq N$ různých hodnot (N_t je kardinalita množiny $X_t \subseteq X$)
- pro příznak x_k můžeme použít prahové hodnoty α_{kn} pro $n=1, \dots, N_{tk}$ ($N_{tk} \leq N_t$), kde α_{kn} je prostřední hodnota mezi dvěma sousedními hodnotami x_k v trénovací množině X

Celkový počet možných dotazů v uzlu t je: $\sum_{k=1}^l N_{tk}$

Podle dělicího kritéria se v uzlu t vybere dotaz, který vede k nejlepšímu dělení množiny X_t v uzlu t .

Dělicí kritérium:

Binární dělení uzlu t :

- vytvoří se dva synovské uzly t_Y a t_N podle odpovědi „Yes“/„No“ na dotaz v uzlu t
- synům přísluší podmnožiny X_{tY} a X_{tN}
- podle stromové struktury (od kořene k listu) každé dělení uzlu t musí vytvořit podmnožiny X_{tY} a X_{tN} , které jsou „více třídově homogenní“ než množina X_t v uzlu t
 - trénovací vzory v každé z nových podmnožin mají „vyšší preference pro danou třídu“
 - trénovací data v X_t jsou stejnoměrněji distribuované mezi různými třídami

Cílem je definovat míru, která určí „nečistotu“ uzlu a rozdělit uzel tak, aby celková „nečistota“ synů byla nižší ve srovnání s „nečistotou“ otce.

Nečistota uzlu se vypočítá pomocí vzorce:

$$I(t) = - \sum_{i=1}^c P(\omega_i|t) \cdot \log_2 P(\omega_i|t)$$

kde $P(\omega_i|t)$ je pravděpodobnost, že vektor z množiny X_t uzlu t spadne do třídy ω_i ($i=1, \dots, c$).

Po rozdělení uzlu do 2 synů:

- N_{tY} vzorů přísluší odpovědi „Yes“ (množina X_{tY})
- N_{tN} vzorů přísluší odpovědi „No“ (množina X_{tN})

Pokles „nečistoty“ uzlu je dán vzorcem:

$$\Delta I(t) = I(t) - \frac{N_{tY}}{N_t} I(t_Y) - \frac{N_{tN}}{N_t} I(t_N)$$

kde $I(t_Y)$ a $I(t_N)$ jsou „nečistoty“ uzlů t_Y a t_N

Cílem je podle dané množiny vytvořit dotaz, který rozdělí uzel tak, aby došlo k největšímu poklesu „nečistoty“.

Ukončovací podmínka :

Kdy ukončit dělení uzlu a prohlásit uzel za list?

- použít prahovou hodnotu T
 - uzel t nebude dělen, když maximum $\Delta I(t)$ přes všechny případné dělení uzlu t je menší než T
- použít velikost X_t
 - uzel t nebude dělen, když kardinalita množiny X_t je dostatečně malá
- použít „čistotu“ množiny X_t

- uzel t nebude dělen, když množina X_t je „čistá“ (tj. ideálně všechny body patří k jediné třídě)

Přiřazení listu ke třídě:

Když uzel prohlásíme za list, dostane „značku“ třídy. Nejčastěji se pro to používá následující pravidlo:

list t dostane značku třídy ω_i když
$$i = \arg \max_j P(\omega_j | t)$$

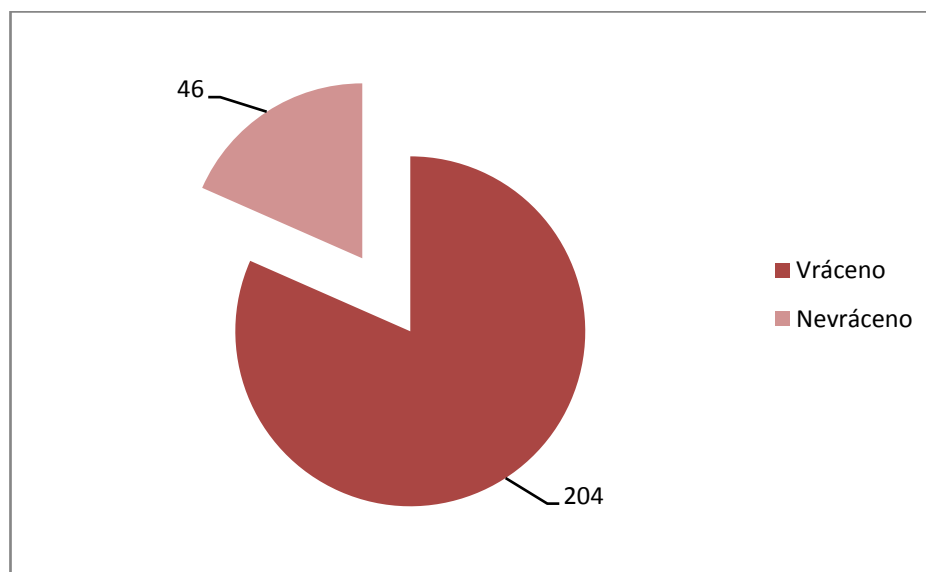
Výhodou klasifikačních stromů je, že pružně postihují vztahy mezi různými typy veličin, nelineární závislosti a interakci proměnných. Ve srovnání s klasickými parametrickými metodami dosahují často srovnatelné přesnosti, ale poskytují přitom daleko přehlednější modely.

Nevýhodou stromů je, že jsou obvykle značně nestabilní: Zhusta pro jedna a tatáž data existuje mnoho různých stromů s přibližně stejnou chybou a při malé změně dat nebo vstupních parametrů se může výsledný strom výrazně změnit [10].

3 METODIKA:

V období od 1.12.2007 do 31.1.2008 jsem prováděla dotazníkový průzkum, týkající se mobilních operátorů v České republice. Celkem jsem oslovila 250 respondentů (z toho 200 osobním kontaktem a 50 prostřednictvím internetu). Míra návratnosti dosáhla 81,6 %, tzn. 204 vrácených dotazníků.

Graf 1: Míra návratnosti



Zdroj: Vlastní tvorba

Dotazník tvořilo 12 otázek a byl rozdělen do dvou oblastí:

- 1) Otázky týkající se problematiky využívání služeb mobilních operátorů
- 2) Otázky týkající se osobních údajů respondenta

Cílovou skupinou byli aktivní uživatelé mobilních telefonů z celé České republiky různých věkových kategorií, s různým povoláním, příjmem a vzděláním.

K vyhodnocení dotazníků jsem použila Excel a programovací prostředí R.

Programovací prostředí R:

R je jazyk a prostředí pro statistické výpočty a grafiku. Je to GNU projekt, který je podobný jako jazyk a prostředí S. R poskytuje širokou statistickou škálu - lineární a nelineární modelové, klasické statistické testy, analýza časových řad, třídění, shlukování,...

Jedna z výhod jazyka R je dobře navržená kvalitní publikační struktura, která obsahuje matematické symboly a rovnice, které jsou velmi potřebné.

V tomto programovacím prostředí jsem ke tvorbě klasifikačních stromů použila následující příkazy:

- a) Pro načtení dat do R: `setwd`, `read.table`
- b) Pro vytvoření klasifikačního stromu: `library`, `rpart`, `plot`, `model.summary`
- c) Pro ořezání klasifikačního stromu: `printrcp`, `modelcp`, `model`, `plot`, `text`

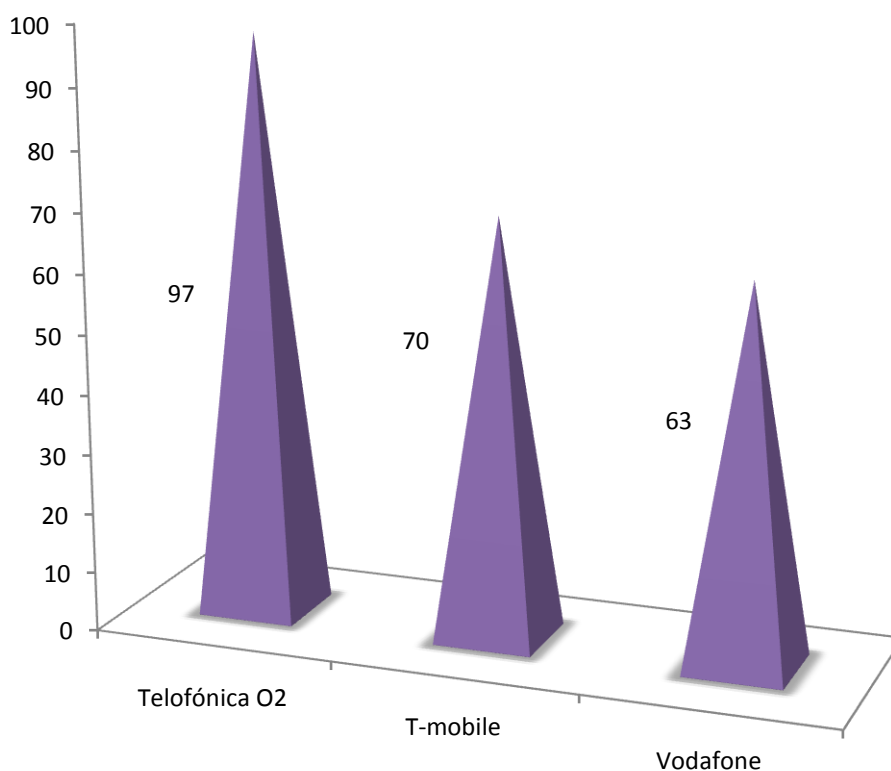
4 VLASTNÍ ZPRACOVÁNÍ:

4.1 Vyhodnocení dotazníků:

Otázku 1 : „Jakého mobilního operátora využíváte?“

Tato otázka je uzavřená a poslední odpověď je filtrová (při zaškrtnutí této odpovědi již dotazovaný dále nedoplňuje). Dotazovaný vybírá z daných možností, které zaškrtně.

Graf 2: „Jakého mobilního operátora využíváte?“



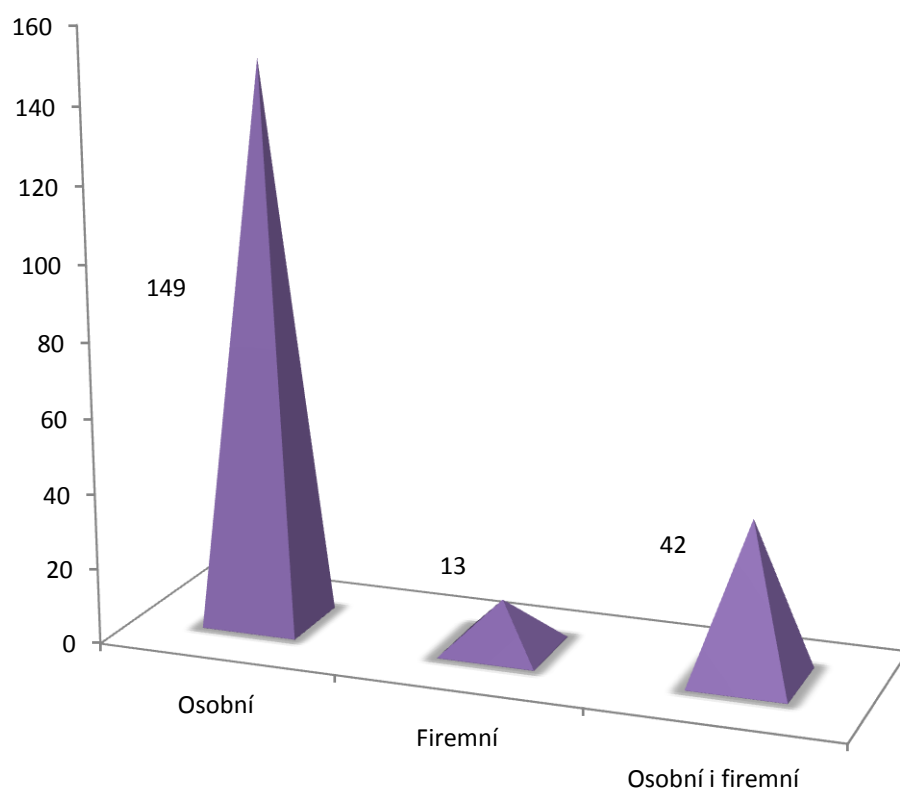
Zdroj: Vlastní tvorba

Z grafu je vidět, že největší okruh dotazovaných, a to téměř 42 %, dává přednost mobilnímu operátoru Telefónica O2, 30 % spotřebitelů dává přednost T-Mobilu a 28 % spotřebitelů preferuje Vodafone.

Otázka 2: „Pro jaké účely?“

Tato otázka je uzavřená a dotazovaný vybírá z daných možností, které zaškrtně.

Graf 3: „Pro jaké účely?“



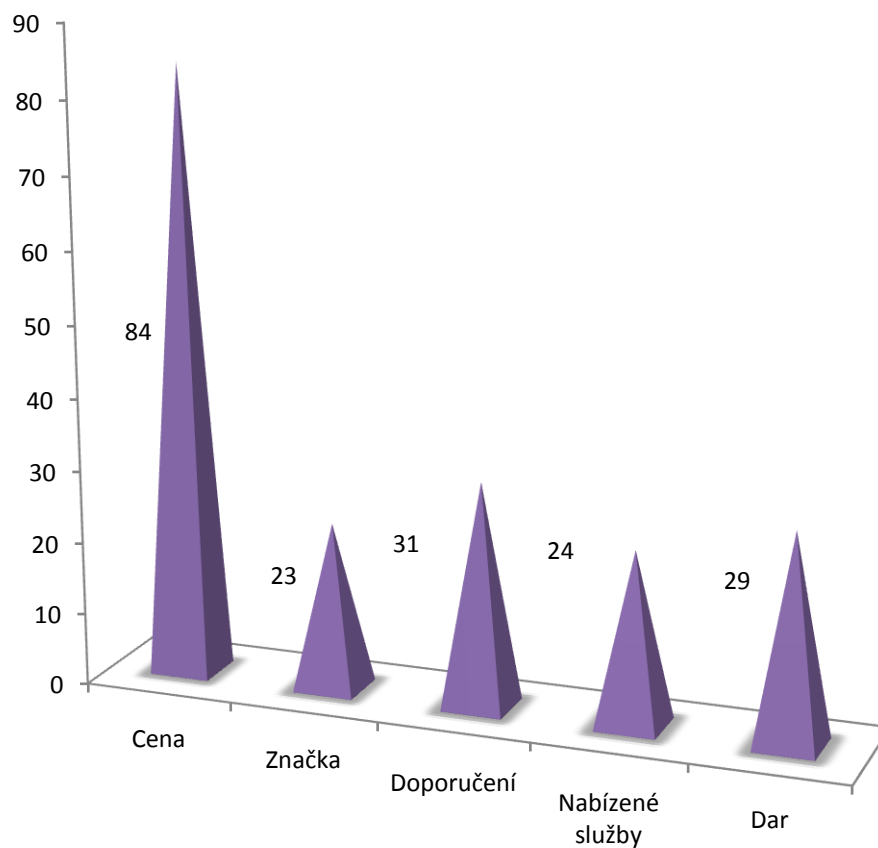
Zdroj: Vlastní tvorba

Nejvíce respondentů využívá mobilního operátora pro osobní účely. Z celého okruhu dotazovaných je to 149 spotřebitelů. Pouze 13 respondentů využívá mobilního operátora jen pro firemní účely a 42 respondentů využívá služeb mobilního operátora jak pro osobní, tak i pro firemní účely.

Otázka 3: „Co pro Vás bylo při výběru mobilního operátora rozhodující?“

Tato otázka je polootevřená, protože dotazovaný má možnosti, jak zaškrtnout odpověď, tak odpověď dopsat.

Graf 4: „Co pro Vás bylo při výběru mobilního operátora rozhodující?“



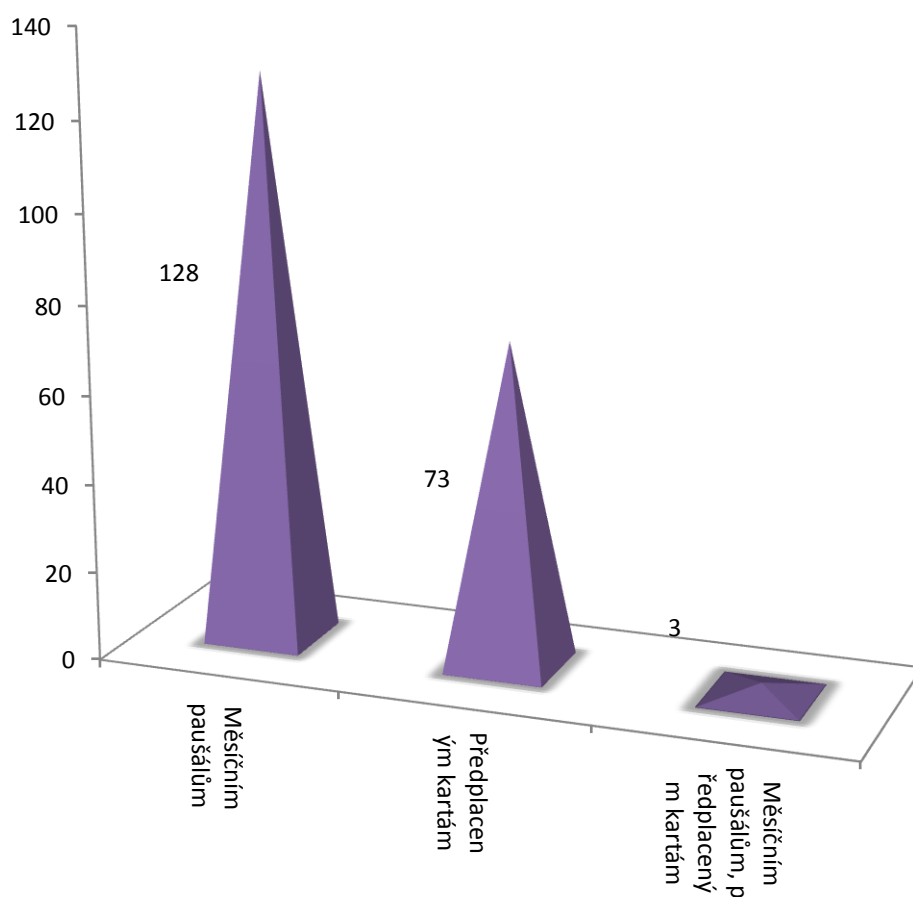
Zdroj: Vlastní tvorba

Výše uvedený graf ukazuje, že čeští spotřebitelé jsou při koupi mobilního operátora nejvíce ovlivněni cenou. Dále si mobilního operátora pořizují na základě doporučení, nebo dostali mobilního operátora darem.

Otázka 4: „Dáváte přednost“

Otázka je uzavřená. Dotazovaný vybírá z otázek, které zaškrtně.

Graf 5: „Dáváte přednost“



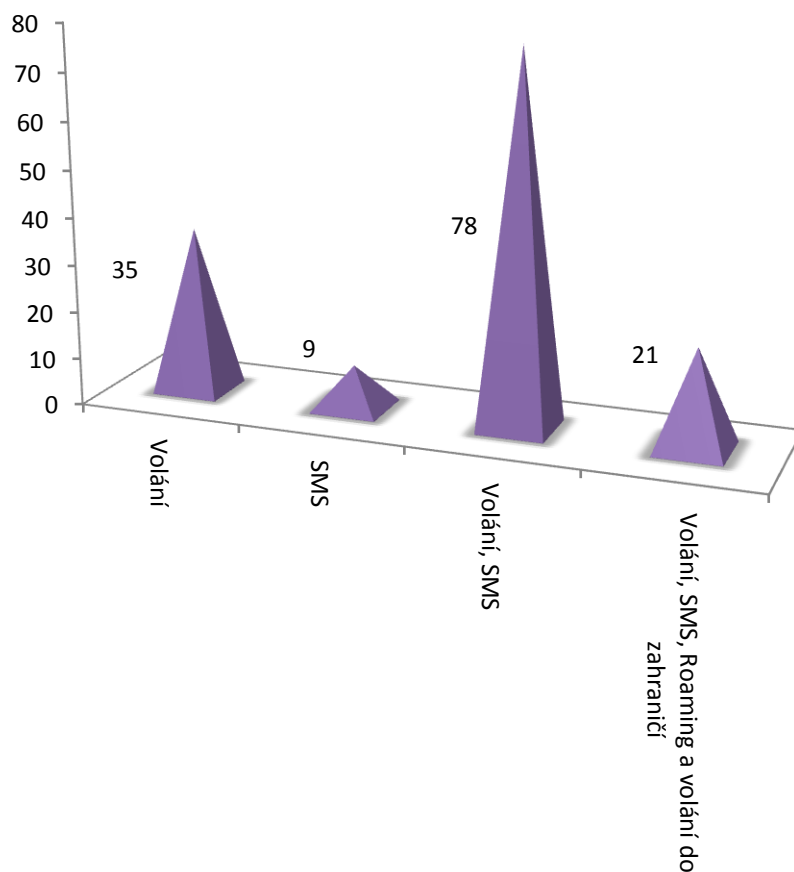
Zdroj: Vlastní tvorba

Z výše uvedeného grafu je patrné, že více než 60 % respondentů dává přednost měsíčním paušálům. U zákazníků mobilních operátorů jsou měsíční paušály čím dál víc oblíbenější. Zákazníci nemusí kontrolovat kredit a mohou volat a posílat SMS zprávy téměř bez omezení.

Otázka 5: „**Které mobilní služby využíváte?**“

Otázka je uzavřená. Dotazovaný vybírá z otázek, které zaškrtně.

Graf 6: „Které mobilní služby využíváte?“



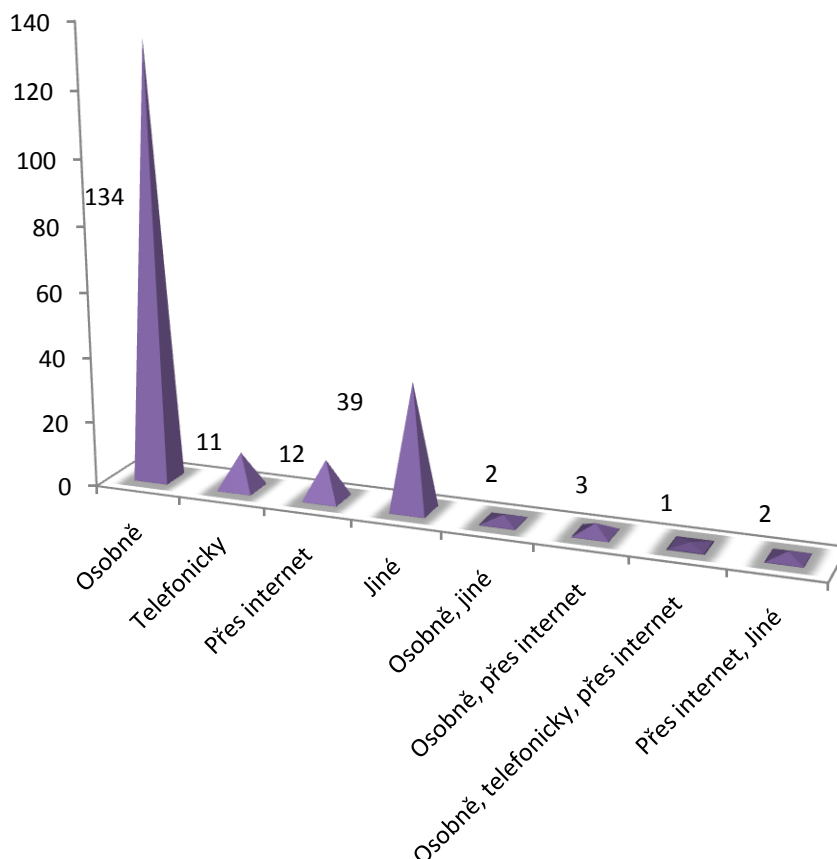
Zdroj: Vlastní tvorba

Jako nejvíce využívané služby zvolili respondenti volání a posílání SMS zpráv. Další nejčastější odpovědi byly, že respondenti jenom telefonují, anebo telefonují, posílají SMS zprávy a využívají roaming a volání do zahraničí.

Otázka 6: „Jakým způsobem jste si mobilního operátora pořídili?“:

Tato otázka je polootevřená, protože dotazovaný má možnosti, jak zaškrtnout odpověď, tak odpověď dopsat.

Graf 7: „Jakým způsobem jste si mobilního operátora pořídili?“



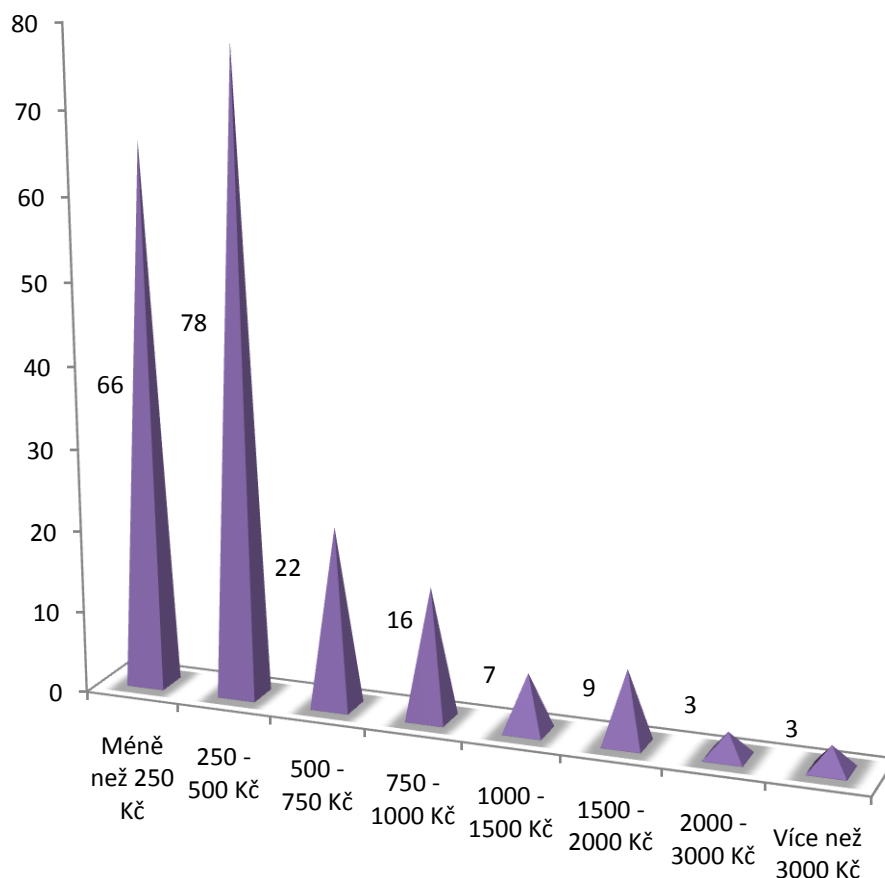
Zdroj: Vlastní tvorba

Graf 7 ukazuje, že si respondenti mobilního operátora pořídili nejčastěji osobně. Je to nejčastější způsob nákupu u nás. I když už v dnešní době využívá hodně lidí k nákupu zboží internet, z výše uvedeného grafu to není patrné. Respondenti to zřejmě nepovažují za příliš bezpečný způsob pořízení zboží. Větší počet respondentů také uvedl, že si mobilního operátora pořídili jiným způsobem, kdy nejčastěji uvedli, že mobilního operátora získali darem nebo si ho obstarali prostřednictvím firmy.

Otázka 7: „**Vaše měsíční útrata za mobilní služby činí**“:

Tato otázka je uzavřená. Dotazovaný vybírá z daných možností, které zaškrtně.

Graf 8: „Vaše měsíční útrata za mobilní služby činí“



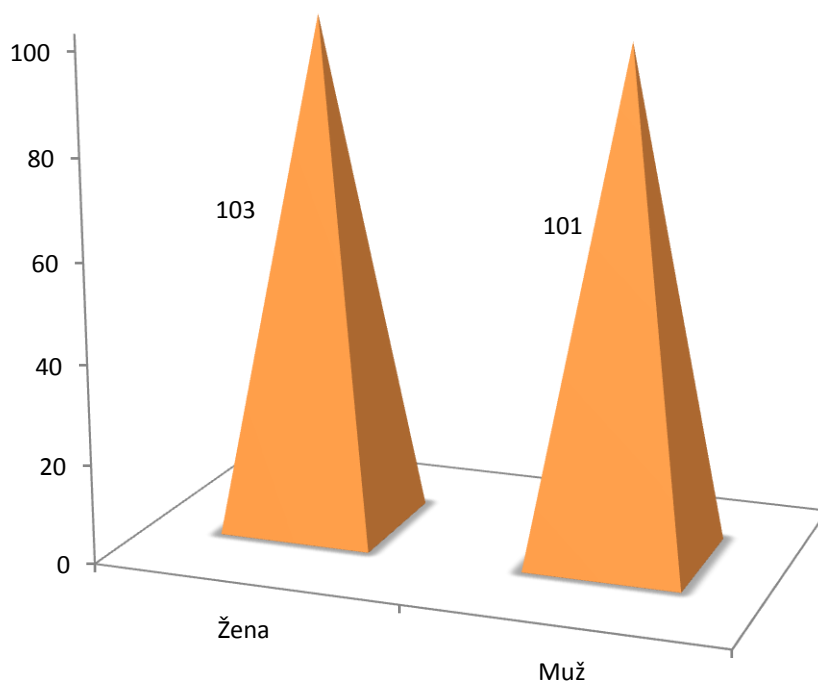
Zdroj: Vlastní tvorba

Z grafu 8 je možné vyčíst, že nejčastější měsíční útraty respondentů jsou menší než 250 Kč a v rozmezí 250 – 500 Kč. Z grafu je také patrné, že pouze šest lidí platí měsíčně za služby mobilních operátorů více než 2000 Kč. Tato otázka může být ovlivněná tím, že mezi dotazovanými byli zaměstnanci společnosti ČEZ, kteří mohou využívat tarify od O2 za bezkonkurenční ceny. Proto jsou jejich útraty za mobilní služby i přes jejich vysoké měsíční příjmy velmi nízké.

Dalších pět otázek se týká osobních charakteristik respondentů. Všechny otázky jsou uzavřené a dotazovaný vybírá z otázek, které zaškrtně.

Otázka 8: „Pohlaví“

Graf 9: „Pohlaví“

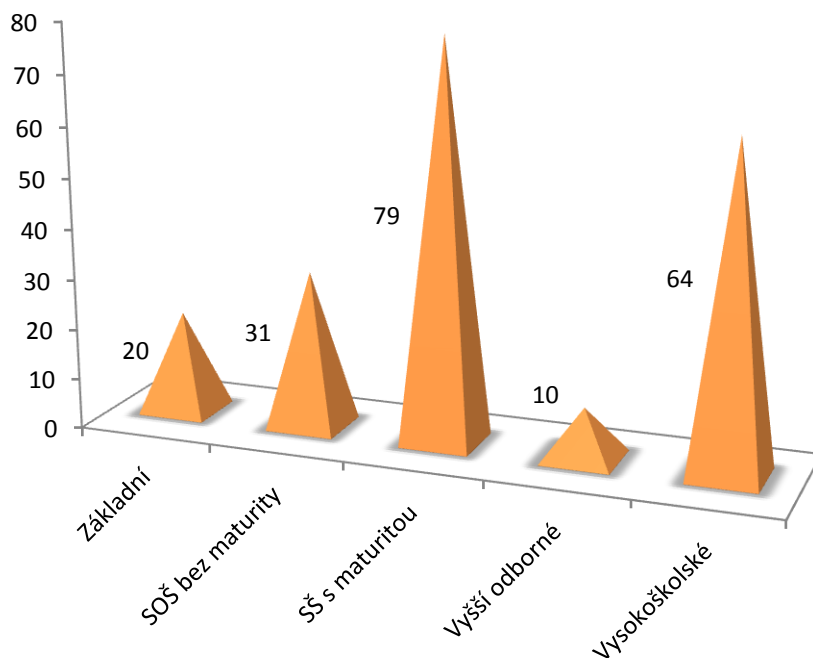


Zdroj: Vlastní tvorba

Poměr dotazovaných žen a mužů byl zhruba vyrovnaný. Dotazník vyplnilo 103 žen a 101 mužů.

Otázka 9: „Nejvyšší ukončené vzdělání“

Graf 10: „Nejvyšší ukončené vzdělání“

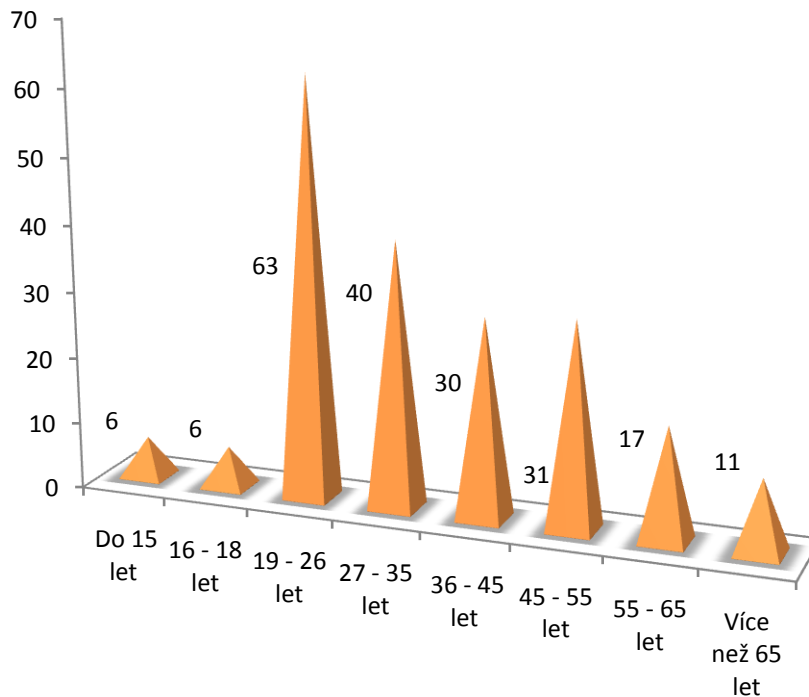


Zdroj: Vlastní tvorba

Ze škály dotazovaných bylo nejvíce středoškoláků (38 %) a vysokoškoláků (32 %). Je to z toho důvodu, že nejvíce byli osloveni studenti a pracovníci na vyšších a vysokých pozicích. Naopak nejméně dotazovaných se přiřadilo do škály základního vzdělání, což byli zejména lidé důchodového věku, kteří neměli možnost studovat a studenti středních nebo základních škol, a také do škály vyššího odborného vzdělání, což může být z toho důvodu, že studenti vyšších odborných škol mají v dnešní době možnost dodělat si vysokoškolské vzdělání a také ke dnes snadnější dostat se na vysokou školu.

Otázka 10: „Věková kategorie“

Graf 11: „Věková kategorie“

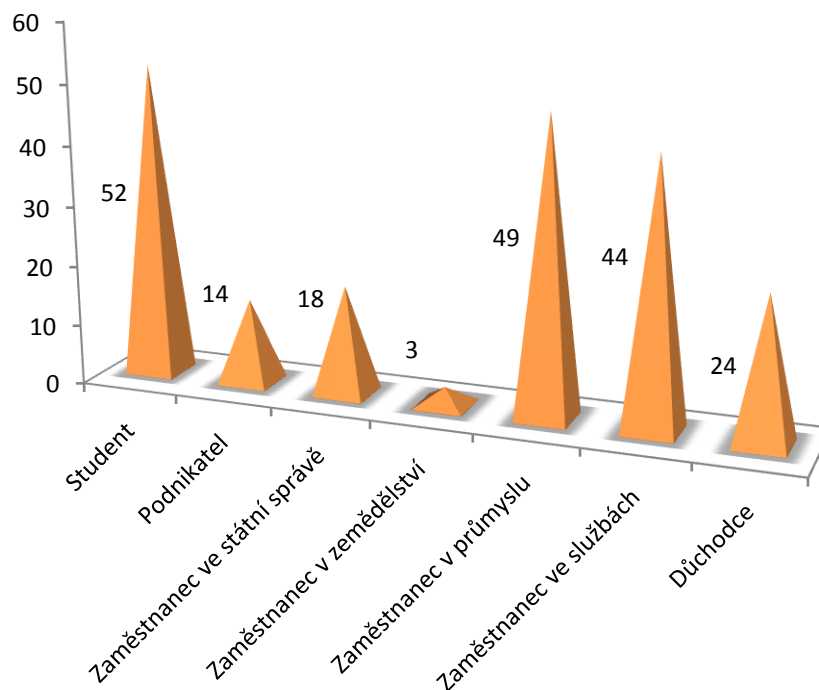


Zdroj: Vlastní tvorba

Výše uvedený graf nám ukazuje, že největší okruh dotazovaných se přiřadil do věkové kategorie 19 – 26 let a dále do věkové kategorie 27 – 35 let a 45 – 55 let. Nejméně oslovených padlo do věkové kategorie do 15ti let a od 16ti – 18ti let. Lidé patřící do věkové kategorie více než 65 let také netvořili příliš velkou škálu dotazovaných. Z toho plyne, že nejvíce byli oslovovali lidé nižších až středních věkových kategorií.

Otázka 11: „Povolání“

Graf 12: „Povolání“

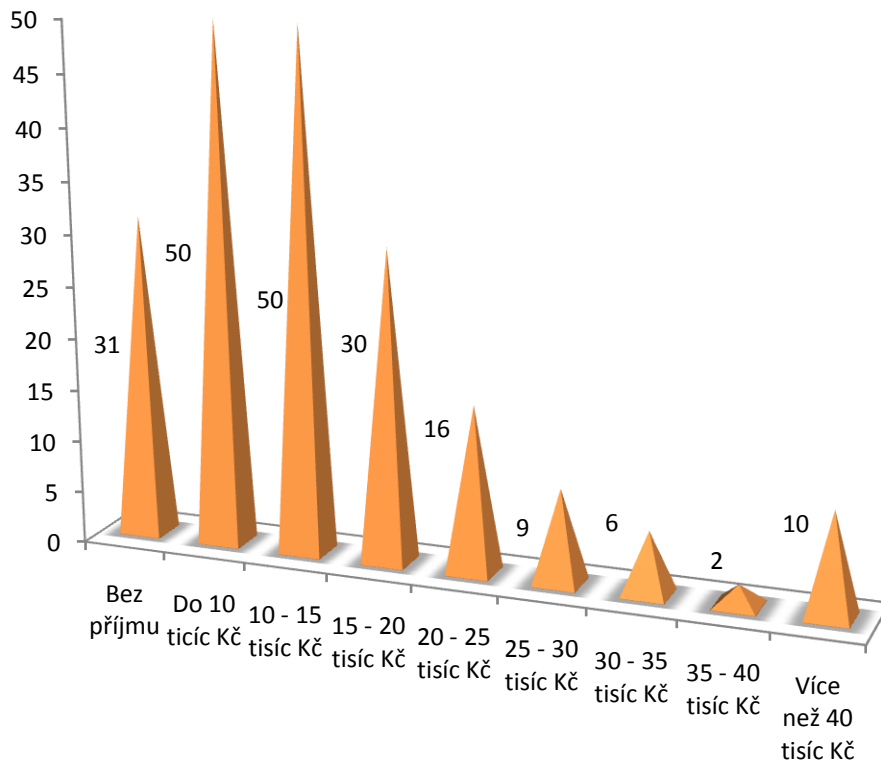


Zdroj: Vlastní tvorba

Graf 12 ukazuje, že nejširší okruh respondentů tvoří studenti a zaměstnanci v průmyslu. Další významnou škálu tvořili respondenti, kteří pracují ve službách. Nejmenší okruh dotazovaných pracuje v oblasti zemědělství a mezi respondenty také není příliš mnoho podnikatelů.

Otázka 12: „Příjmová kategorie“

Graf 13: „Příjmová kategorie“



Zdroj: Vlastní tvorba

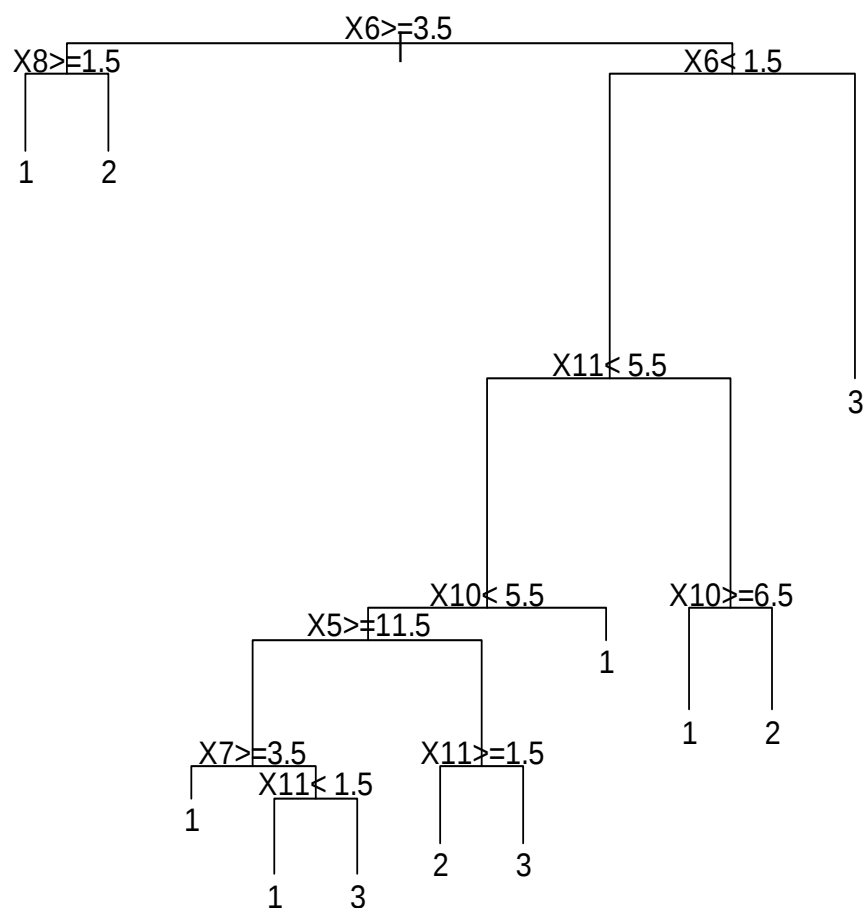
Nejvíce respondentů se přiřadilo do příjmové kategorie do 10. tisíc Kč a do kategorie 10 – 15 tisíc Kč. Větší kategorii také tvořil okruh dotazovaných spadajících do kategorie bez příjmu, kterou tvořili jen studenti, a také okruh dotazovaných z kategorie 15 – 20 tisíc Kč. 10 dotazovaných má příjem vyšší než 40 tisíc Kč. Jsou to především pracovníci na vysokých pozicích ze společnosti ČEZ a někteří podnikatelé.

4.2 Sestrojení klasifikačních stromů:

Jakými faktory jsou při výběru mobilního operátora mobilní operátoři nejvíce ovlivněni?

Odpověď na tuto otázku ukazuje následující neořezaný a ořezaný klasifikační strom.

Graf 14: Kompletní klasifikační strom vygenerovaný v rpart



Zdroj: Vlastní tvorba

Graf 14 je kompletní neořezaný klasifikační strom vygenerovaný v rpart. Klasifikuje respondenty podle toho, co je nejvíce ovlivňuje při výběru mobilního operátora: Je to otázka 6 („Jakým způsobem jste si mobilního operátora pořídili?“), otázka 8 („Pohlaví“), otázka 11 („Povolání“), otázka 10 („Věková kategorie“), otázka 5 („Jaké mobilní služby využíváte?“) a otázka 7 („Vaše měsíční útrata za služby činí“). Každý dotaz odpovídá dělení uzlu do 2 synů podle odpovědi ano/ ne na dotaz v uzlu. Jestliže prohlásíme kořen za list, dostane přidělenou značku třídy. Kořenovým uzlem tohoto stromu je uzel X6 („Jakým způsobem jste si mobilního operátora pořídili?“). Strom má dále devět větví a jedenáct terminálních listů.

Dále budu strom ořezávat:

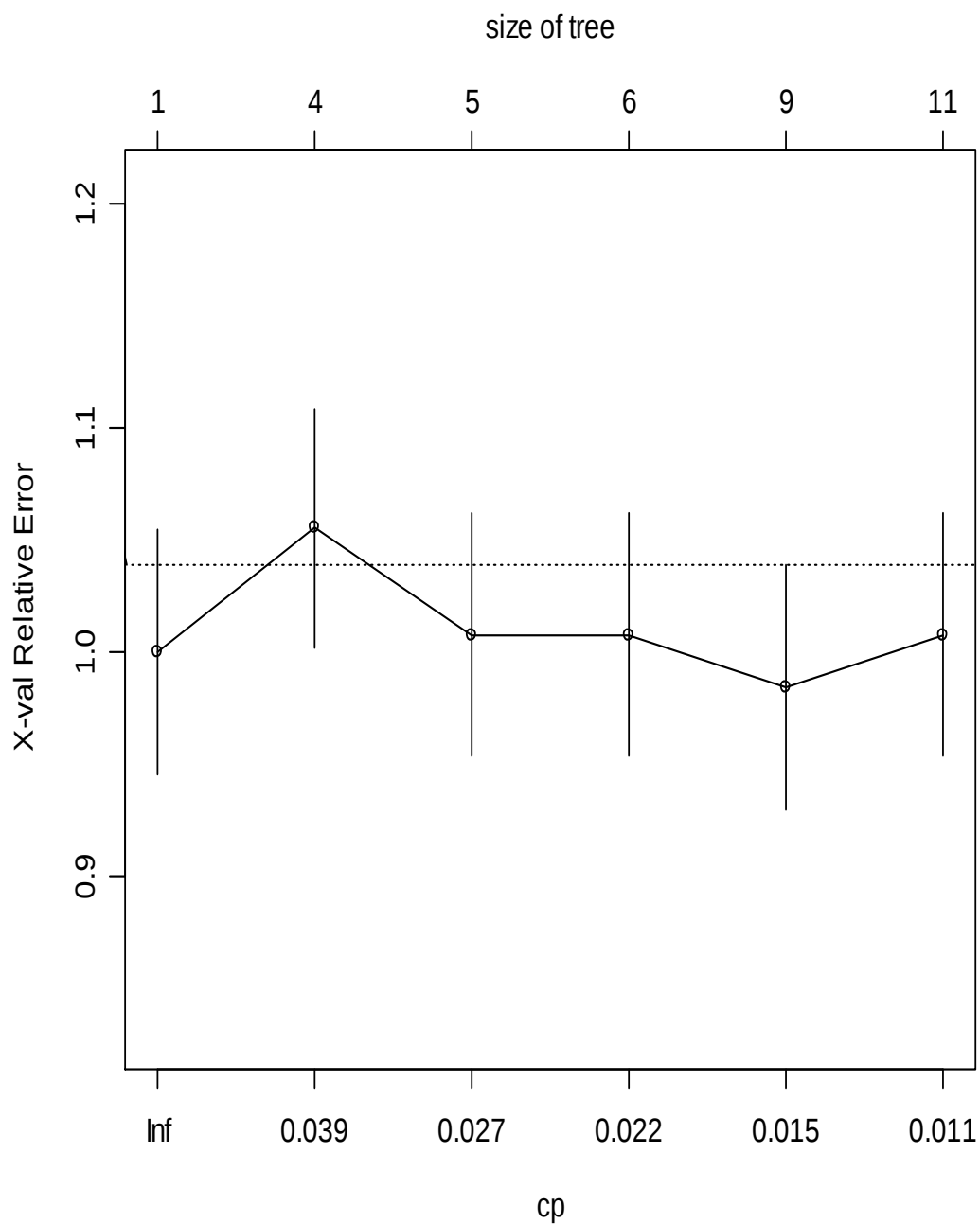
Tabulka 1: Parametr složitosti

	CP	nsplit	rel error	xerror	xstd
1	0.047244	0	1.00000	1.00000	0.054517
2	0.031496	3	0.83465	1.05512	0.053393
3	0.023622	4	0.80315	1.00787	0.054374
4	0.019685	5	0.77953	0.98425	0.054374
5	0.011811	8	0.71654	0.98425	0.054783
6	0.010000	10	0.69291	1.00787	0.054374

Zdroj: Vlastní tvorba

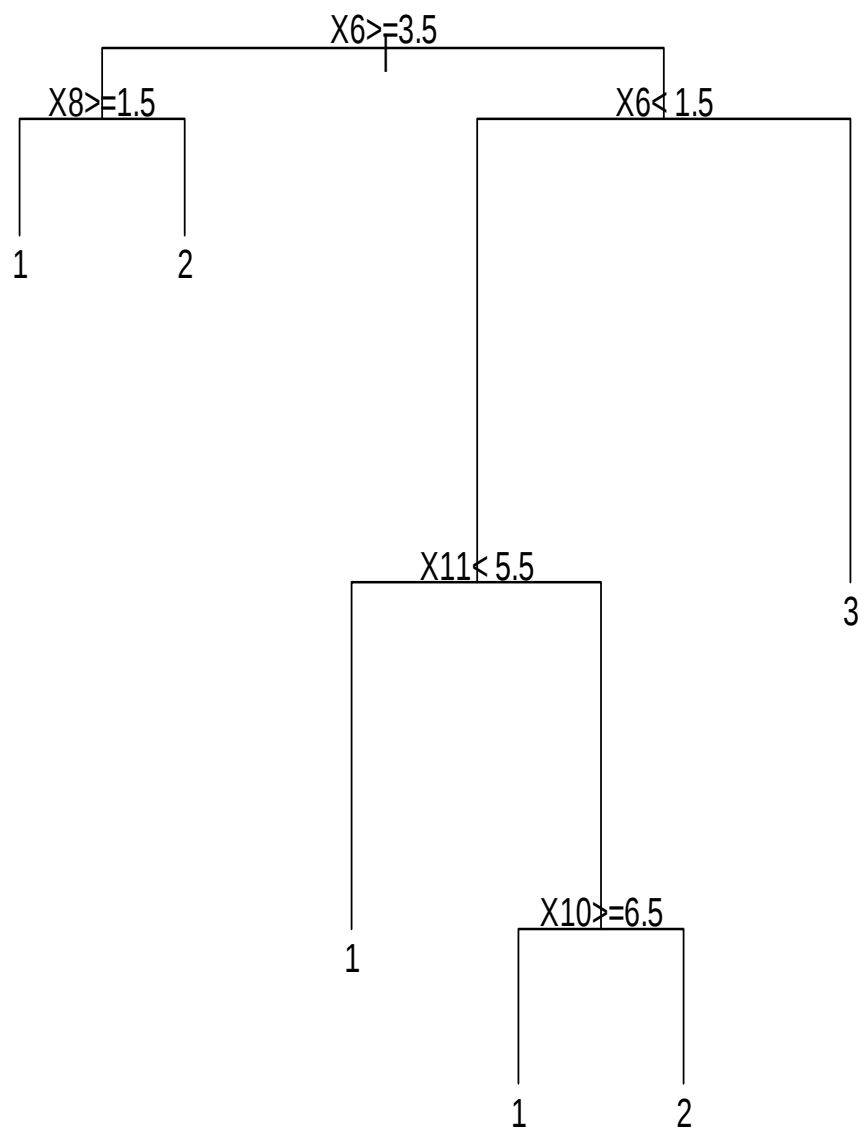
Výše uvedená tabulka ukazuje parametr složitosti (CP). Hlavním důvodem pro používání CP je „vykalkulovat“ zlepšení, vyplývající z dalšího rozvětvení v datech. To také měří přesnost ztracenou odstraněním jednoho nebo více koncových uzlů [11]. Strom s jedním koncovým uzlem snižuje chybu o 47 %; strom se dvěma koncovými uzly ji dále zlepšuje o 31 %. Standardně, rpart zastaví růst stromu, když je $CP = 0,01$.

Graf 15: 1 SE pravidlo a CP



Optimálně prořezaný strom vybereme 1 SE pravidlem na základě parametru složitosti (CP), jak je vidět na výše uvedeném grafu.

Graf 16: Ořezaný klasifikační strom vygenerovaný v rpart

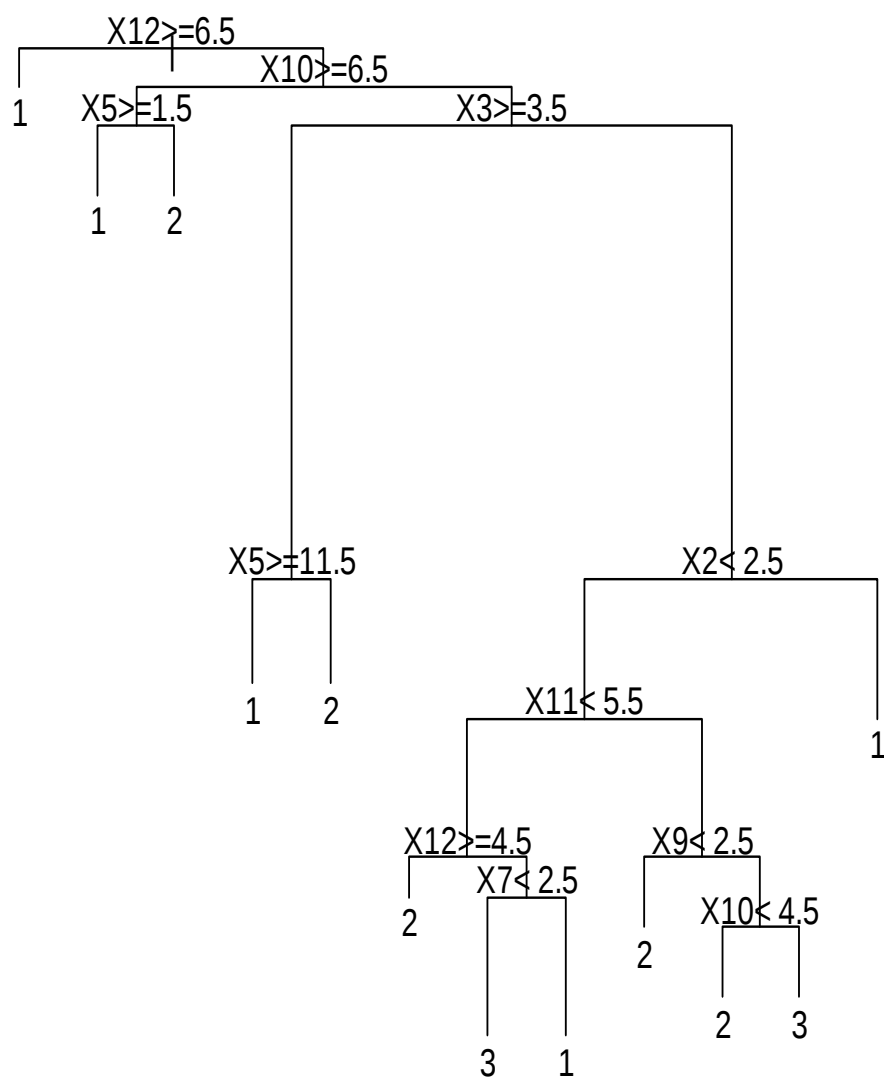


Zdroj: Vlastní tvorba

Na grafu 14 je už ořezaný klasifikační strom vygenerovaný v rpart. Ořezáním stromu se z něho odstranily otázky 5 a 7. Respondenti se tedy podle tohoto klasifikačního stromu rozhodují podle toho, jakým způsobem si mobilního operátora pořídili, podle pohlaví, povolání a věku. Kořenovým uzlem tohoto stromu je uzel X6 („Jakým způsobem jste si mobilního operátora pořídili?“). Dále má strom čtyři větve a šest terminálních listů.

Protože data z dotazníků byla kódována v Excelu a podle mého názoru otázka číslo 6, která je kořenovým uzlem a zní: „Jakým způsobem jste si mobilního operátora pořídili“, nebude tím nejdůležitějším faktorem, na základě kterého se spotřebitelé rozhodují. Proto jsem se rozhodla sestavit ještě jeden klasifikační strom, ve kterém tuto otázku vynechám.

Graf 17: Klasifikační strom vygenerovaný v rpart bez faktoru X6



Zdroj: Vlastní tvorba

Graf 17 je opět kompletní neořezaný klasifikační strom vygenerovaný v rpart. Klasifikuje respondenty podle toho, co je nejvíce ovlivňuje při výběru mobilního operátora, ovšem už bez otázky č. 6: Jakým způsobem jste si mobilního operátora pořídili. Díky této redukci v tomto stromu přibýly další faktory, které respondenty při jejich výběru ovlivňují. Je to otázka 12 („Váš čistý měsíční příjem“), otázka 3 („Co pro Vás bylo při výběru mobilního operátora rozhodující?“), otázka 2 („Pro jaké účely?“) a otázka 9 („Nejvyšší ukončené vzdělání“). Naopak se v tomto stromu neobjevila otázka 8 („Pohlaví“). Kořenovým uzlem je uzel X12 a strom má dále deset větví a dvanáct terminálních listů.

Dále budu strom ořezávat.

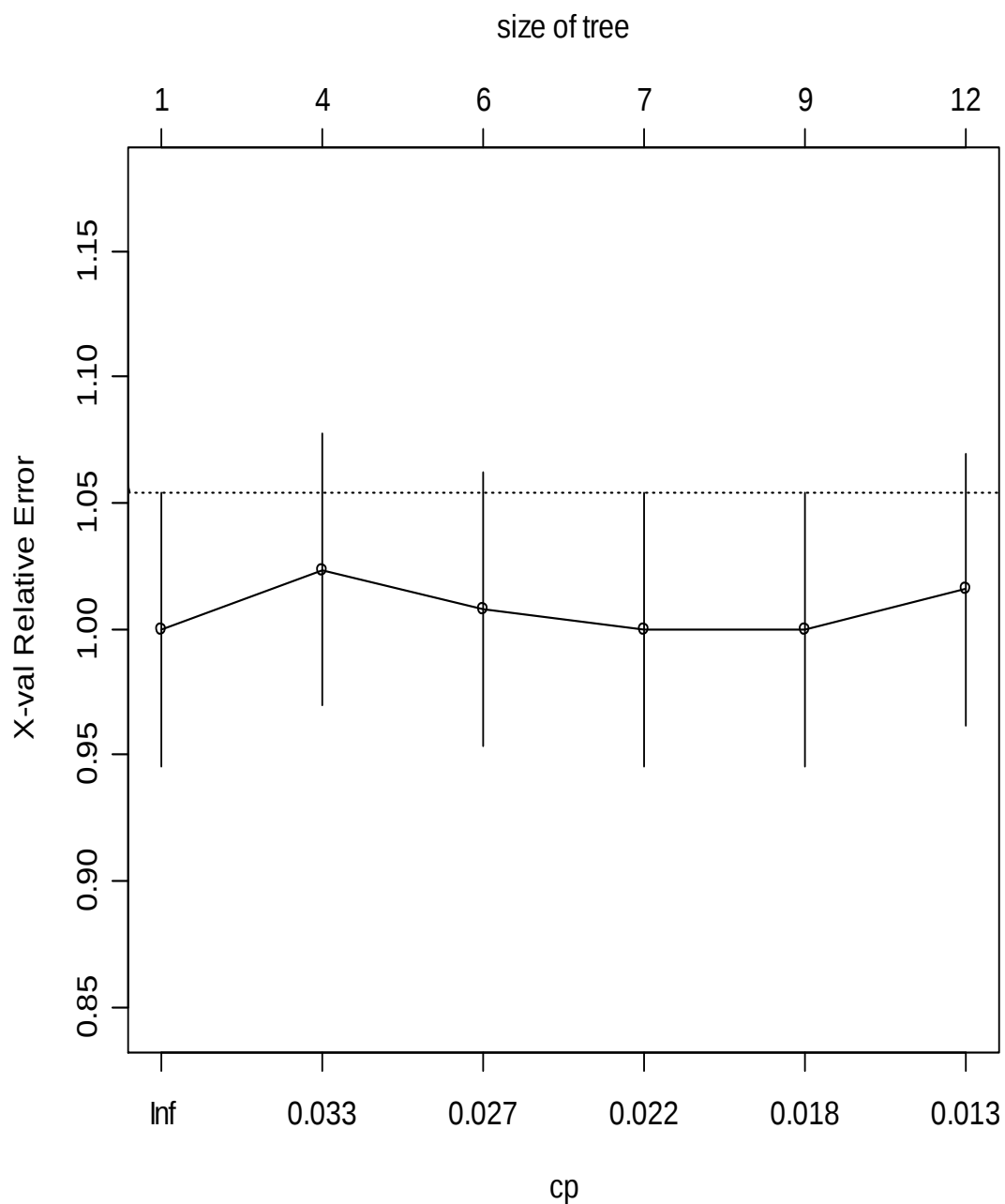
Tabulka 2: Parametr složitosti

	CP	nsplit	rel error	xerror	xstd
1	0,034121	0	1,00000	1,0000	0,054517
2	0,031496	3	0,89764	1,0236	0,054072
3	0,023622	5	0,83465	1,0079	0,054072
4	0,019658	6	0,81102	1,0000	0,054517
5	0,015748	8	0,77165	1,0000	0,054517
6	0,010000	11	0,72441	1,0157	0,054226

Zdroj: Vlastní tvorba

Tato tabulka ukazuje opět parametr složitosti (CP). Můžeme v ní vyčíst, že strom s jedním koncovým uzlem snižuje chybu o 34 %; strom se dvěma koncovými uzly ji dále zlepšuje o 31 %.

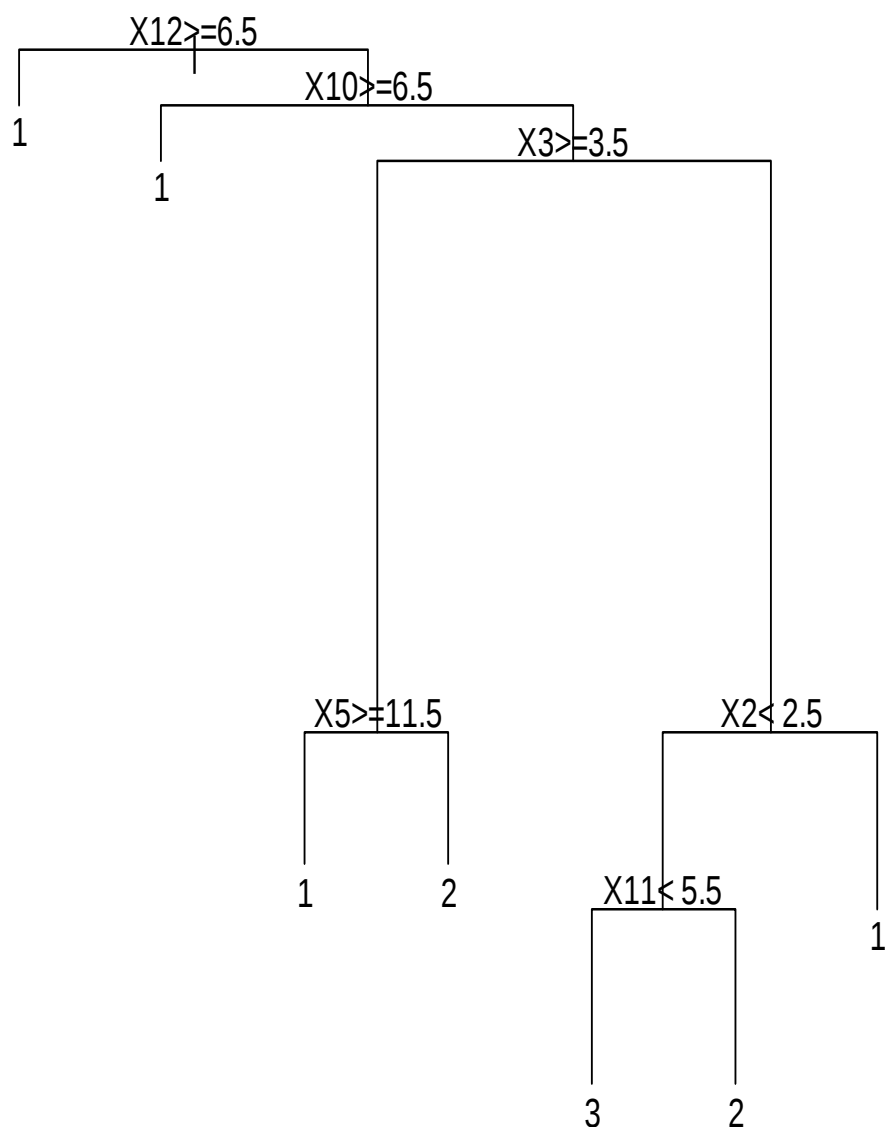
Graf 18: 1SE pravidlo a CP



Zdroj: Vlastní tvorba

Optimálně prořezaný strom vybereme 1 SE pravidlem na základě parametru složitosti (CP), jak je vidět na výše uvedeném grafu.

Graf 19: Ořezaný klasifikační strom vygenerovaný v rpart bez faktoru X6



Zdroj: Vlastní tvorba

Graf 19 představuje již prořezaný klasifikační strom, který se zredukoval oproti neořezanému stromu o uzle X7 a X9. Z tohoto klasifikačního stromu tedy vyplývá, že spotřebitelé jsou při svém výběru mobilního operátora nejvíce ovlivněni výší čistého měsíčního příjmu, věkem, tím co pro ně bylo při výběru mobilního operátora rozhodující, tím jaké mobilní služby využívají, pro jaké účely a také tím jaké povolání vykonávají. Kořenovým uzlem je uzel X12, tedy čistý příjem spotřebitelů. Strom dále tvoří pět větví a sedm terminálních listů.

Větvení:

a) Uzel X12 („Příjmová kategorie):

Tento uzel je kořenovým uzlem klasifikačního stromu. Větví se do dvou synů podle odpovědi „ano/ne“ na dotaz, zda je $X12 \geq 6,5 \Leftrightarrow$ je-li měsíční příjem respondentů vyšší nebo roven 25 tisícům Kč půjde doleva a stává se terminálním listem s třídním označením 1, a nebude tedy už dále větven. V opačném případě, jestliže je měsíční příjem respondentů nižší než 25 tisíc Kč, jde doprava do uzlu X10, který je synem uzlu X12.

b) Uzel X10 („Věková kategorie“):

Uzel, vyjadřující věkovou kategorii respondentů, je synem předchozího uzlu a zároveň otcem uzlu X3. Větví se do dvou synů na základě odpovědi na dotaz, zda je $X10 \geq 6,5 \Leftrightarrow$ je-li odpověď na tuto otázku kladná a respondentům je tedy 46 nebo více let, větví se uzel doleva a stává se terminálním listem s třídním označením 1. Jestliže jsou respondenti mladší, půjde doprava do synovského uzlu X3.

c) Uzel X3 („Co pro Vás bylo při výběru mobilního operátora rozhodující?“):

Tento uzel vyjadřuje faktory, které byly při výběru mobilního operátora pro respondenty rozhodující. Je synovským uzlem uzlu, který vyjadřuje věkovou kategorii a otcem uzlů X5 a X2. Větví se na základě odpovědi „ano/ne“ na dotaz, zda je $X3 \geq 3,5 \Leftrightarrow$ jestliže se respondenti rozhodují

pro mobilního operátora na základě doporučení, či že ho získali darem, půjde doleva do uzlu X5. V opačném případě, jestliže respondenty při jejich výběru nejvíce ovlivňuje cena a značka, půjde uzel doprava do uzlu X2.

d) Uzel X5 („Které mobilní služby nejvíce využíváte?“)

Pod tímto uzlem jsou skryté všechny možné mobilní služby, které mohou zákazníci mobilních operátorů využívat. Větví se na základě dotazu „je $X5 \geq 11,5$?“ $\Leftrightarrow$ v případě, že je odpověď „ano“ a respondenti využívají nejvíce volání a posílání SMS zpráv, bude se tento uzel větvit doleva a stane se terminálním listem s třídním označením 1 a dále se nebude větvit. V případě odpovědi „ne“, využívají respondenti i jiné služby, půjde uzel doprava a opět se stane terminálním listem s třídním označením 2 a dále se už nebude větvit.

e) Uzel X2 („Pro jaké účely“)

To, zda respondenti využívají mobilního operátora pro osobní či firemní účely, vyjadřuje uzel X2. Je synem uzlu X3 a zároveň otcem uzlu X11. Větví se na základě dotazu, je-li $X2 < 2,5$. $\Leftrightarrow$ V případě kladné odpovědi využívají respondenti mobilního operátora pouze pro osobní či firemní účely, a tento uzel jde doleva do uzlu X11. V případě záporné odpovědi budou respondenti využívat mobilního operátora pro osobní a zároveň i pro firemní účely, a uzel se bude větvit doprava a stane se terminálním listem s třídním označením 1 a dále se už nebude větvit.

f) Uzel X11 („Povolání“):

Tento uzel skrývá jednotlivá povolání respondentů. Je synem uzlu X2 a větví se podle odpovědi na dotaz, zda je $X11 < 5,5$ $\Leftrightarrow$ v případě kladné odpovědi, jsou respondenti studenti, podnikatelé anebo jsou zaměstnanci v zemědělství či průmyslu, půjde uzel doleva a stane se terminálním listem s třídním označením 3. V případě záporné odpovědi jsou respondenti zaměstnaný

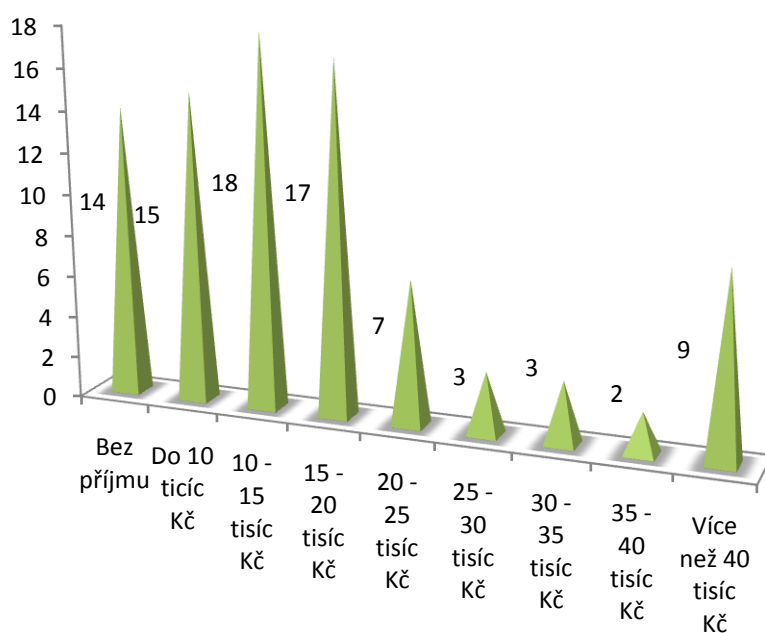
v oblasti služeb, tudíž uzel půjde doprava a stane se terminálním listem s třídním označením 2.

4.3 Diskuse:

Respondenti jsou tedy při výběru mobilního operátora ovlivněni:

- a) Jejich **čistým měsíčním příjmem**. Mobilní operátoři nabízejí mnohdy své mobilní služby za velmi odlišné ceny. Na grafech níže je zobrazeno, kolik lidí z každé příjmové kategorie využívá jednotlivé operátory.

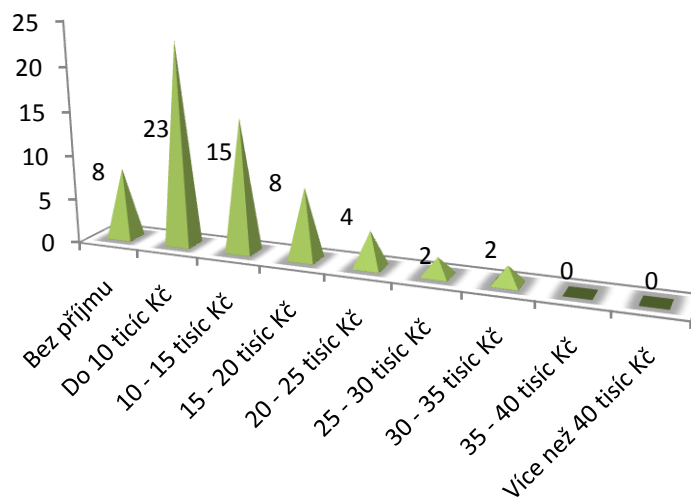
Graf 20: Čistý měsíční příjem zákazníků Telefonie O2



Zdroj: Vlastní tvorba

Tomuto mobilnímu operátorovi dává přednost především okruh respondentů spadající do příjmové kategorie 10 – 15 tisíc Kč a 15 – 20 tisíc Kč. Je také zajímavé, že z deseti respondentů, kteří se zařadili do nejvyšší příjmové kategorie, využívá právě tohoto mobilního operátora. Do této kategorie se zařadili zejména zaměstnanci společnosti ČEZ.

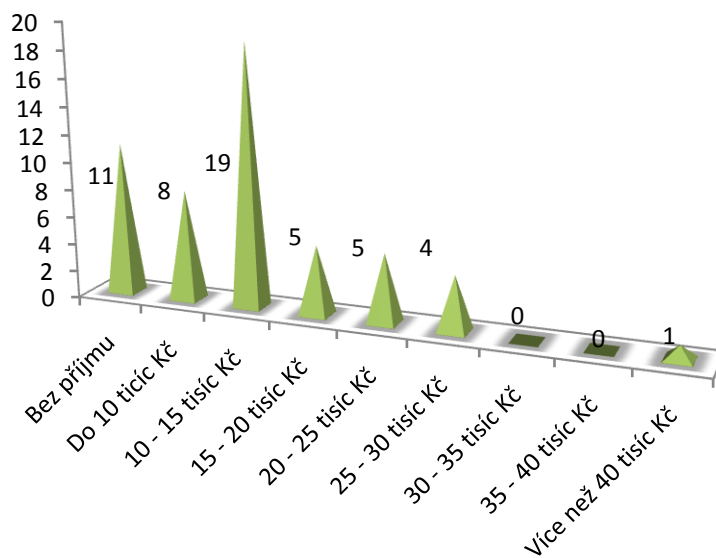
Graf 21: Čistý měsíční příjem zákazníků T- Mobilu



Zdroj: Vlastní tvorba

Tento graf znázorňuje, že tohoto mobilního operátora jednoznačně preferuje okruh respondentů, kteří patří do příjmové kategorie do 10 tisíc Kč. Respondenti s vyššími příjmy T-Mobile příliš nepreferují.

Graf 22: Čistý měsíční příjem zákazníků Vodafonu



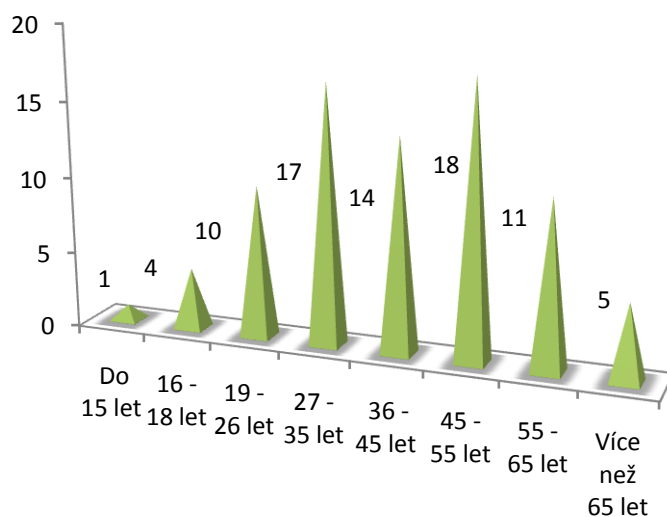
Zdroj: Vlastní tvorba

Na grafu 22 je vidět, že Vodafone využívá zejména skupina respondentů, jejichž měsíční příjem je od 10 – 15 tisíc Kč. Vodafone také dává přednost skupiny respondentů, kteří jsou bez pravidelného měsíčního příjmu. Tuto skupiny tvořili jenom studenti. Ani tohoto operátora nepreferují lidé s vyššími příjmy.

Respondenti s vyššími příjmy preferují spíše Telefónicu O2, zatímco Vodafone a T-Mobile si pořizují respondenti s průměrnými příjmy.

b) Věkovou kategorií

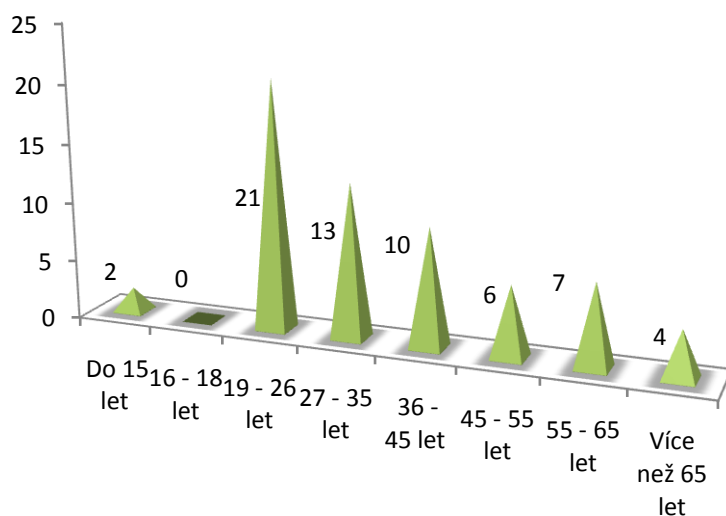
Graf 23: Věková kategorie zákazníků Telefónici O2



Zdroj: Vlastní tvorba

Telefónica O2 dává přednost okruh dotazovaných spadajících do střední věkové kategorie (v rozmezí od 27 do 55 let).

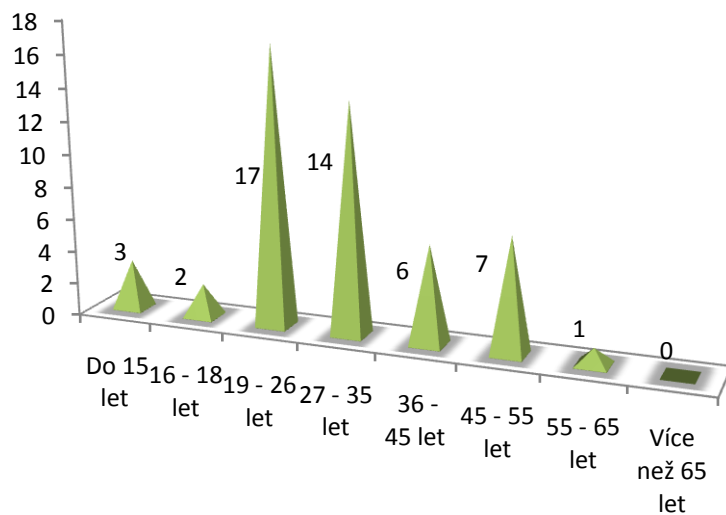
Graf 24: Věková kategorie zákazníků T-Mobilu



Zdroj: Vlastní tvorba

T-Mobile je nejvíce oblíben u věkové kategorie 19 – 26 let.

Graf 25: Věková kategorie zákazníků Vodafonu



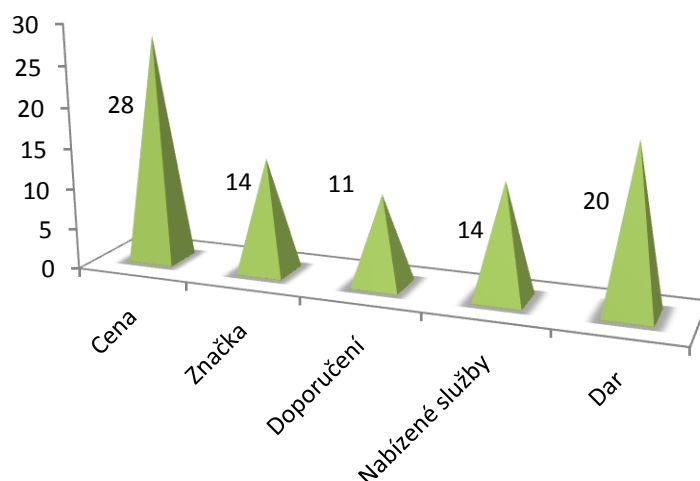
Zdroj: Vlastní tvorba

Vodafone využívá nejvíce, podobně jako u T-Mobilu, skupina dotazovaných z věkové kategorie od 19 – 36 let a také dotazovaní z kategorie 27 – 36 let.

Telefónicu O2 tedy preferuje spíše skupina respondentů středního věku, zatímco Vodafone a T-Mobile využívá spíše mladší věková kategorie respondentů.

c) Respondenti dále volili mobilního operátora podle toho, co pro ně bylo při koupi rozhodující.

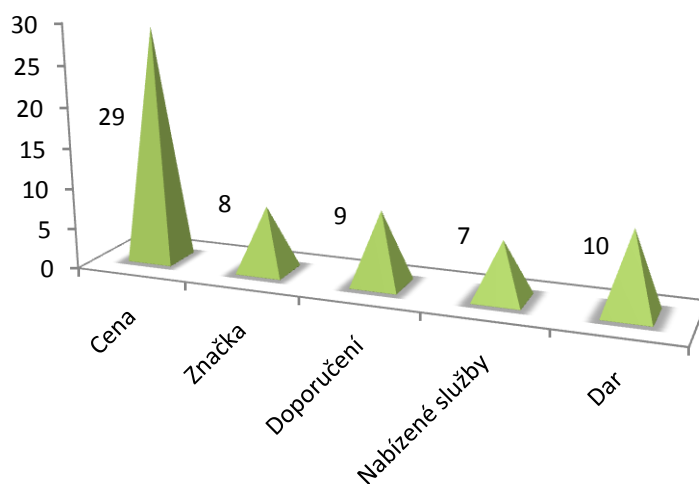
Graf 26: Rozhodující faktory zákazníků Telefónici O2



Zdroj: Vlastní tvorba

Při výběru tohoto mobilního operátora hrála u respondentů největší roli cena a také dar. Nejméně potom respondenti dali na doporučení někoho jiného.

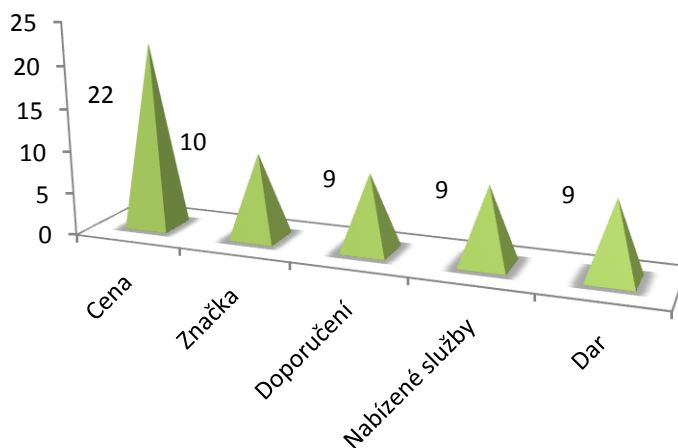
Graf 27: Rozhodující faktory zákazníků T-Mobilu



Zdroj: Vlastní tvorba

Pořízení tohoto mobilního operátora bylo ovlivněno jednoznačně cenou, nejméně potom tím, jaké tento operátor nabízí služby.

Graf 28: Rozhodující faktory zákazníků Vodafone



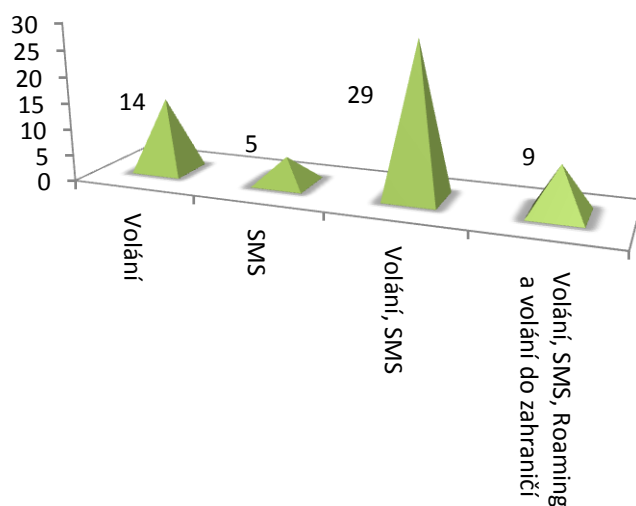
Zdroj: Vlastní tvorba

Podobně jako u T-Mobilu, respondenty při pořízení tohoto operátora nejvíce ovlivnila cena. Další faktory byly na srovnatelné úrovni.

Při rozhodování o nákupu mobilního operátora hrála jak u Telefónici O2, tak u T-Mobilu i Vodafoneu hlavní roli především cena.

d) Dále byli respondenti ovlivněni tím, jaké mobilní služby využívají.

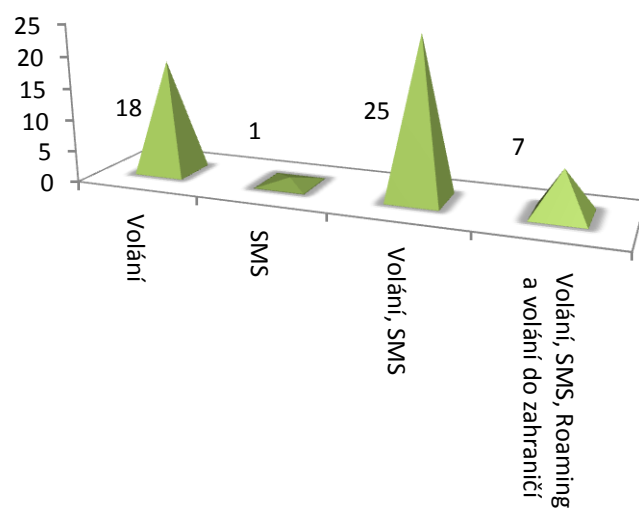
Graf 29: Využívané služby zákazníky Telefónici O2



Zdroj: Vlastní tvorba

Respondenti, kteří jsou zákazníci Telefónici O2, nejčastěji telefonují a posílají SMS zprávy. Pouze v pěti případech posílají pouze textové zprávy.

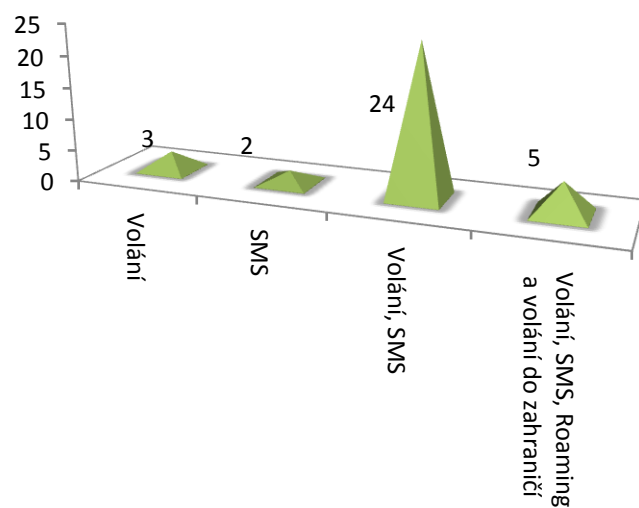
Graf 30: Využívané služby zákazníky T-Mobilu



Zdroj: Vlastní tvorba

Podobně jako u předchozího operátora i zákazníci T-Mobilu nejvíce využívají služeb volání a posílání SMS zpráv a nejméně potom posílají jen SMS zprávy.

Graf 31: Využívané služby zákazníky Vodafone



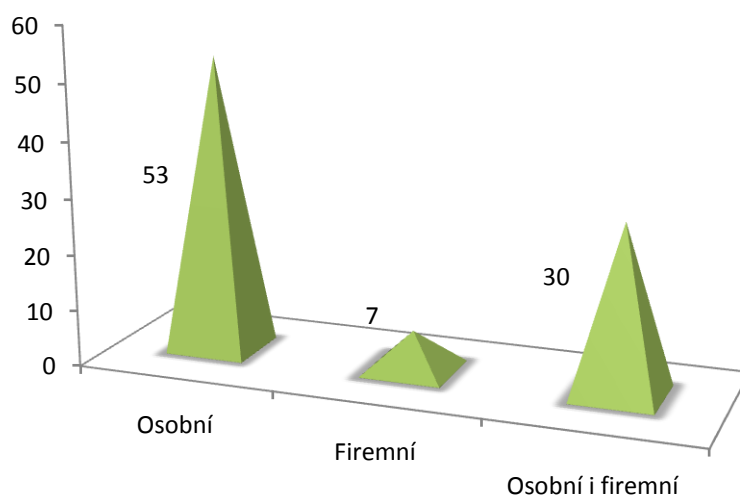
Zdroj: Vlastní tvorba

To samé, jako jsem zmínila u Telefónici O2 a u T-Mobilu, platí i pro tohoto operátora.

U všech mobilních operátorů respondenti využívají nejvíce služeb volání společně s posíláním textových zpráv, nejméně potom posílají jenom textové zprávy.

e) Dále respondenti volí mezi operátory podle toho, zda ho budou využívat pro osobní či firemní účely.

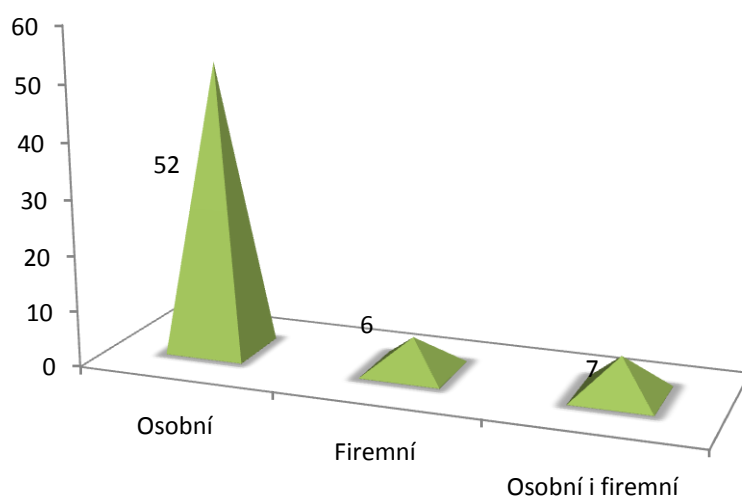
Graf 32: Využívání Telefonici O2 pro osobní či firemní účely



Zdroj: Vlastní tvorba

Výše uvedený graf ukazuje, že skupina respondentů, která zahrnuje 53 osob, využívá Telefonici O2 pouze pro osobní účely. Velká skupina respondentů tohoto operátora využívá ovšem i pro firemní účely.

Graf 33: Využívání T-Mobilu pro osobní či firemní účely

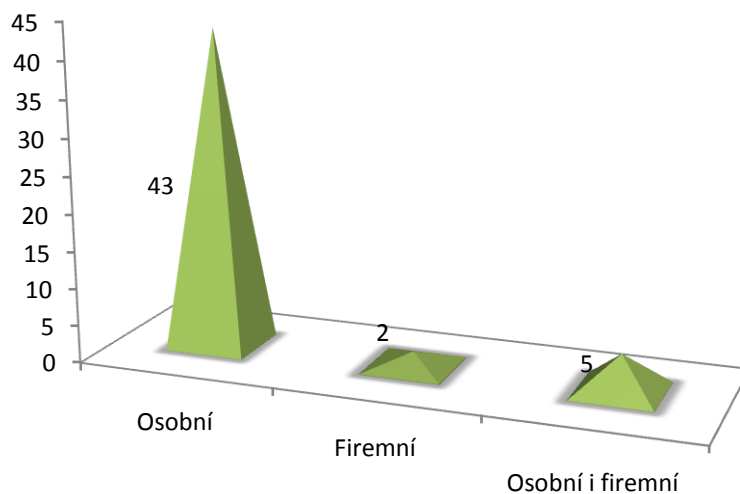


Zdroj: Vlastní tvorba

Největší skupina respondentů využívá T-Mobile pouze pro osobní účely. Pouze zanedbatelné procento T-Mobile využívá i pro firemní účely.

Největší počet zákazníků, kteří využívají mobilního operátora i pro firemní účely, má Telefónica O2. U ostatních mobilních operátorů je tento počet zanedbatelný.

Graf 34: Využívání Vodafone pro osobní či firemní účely

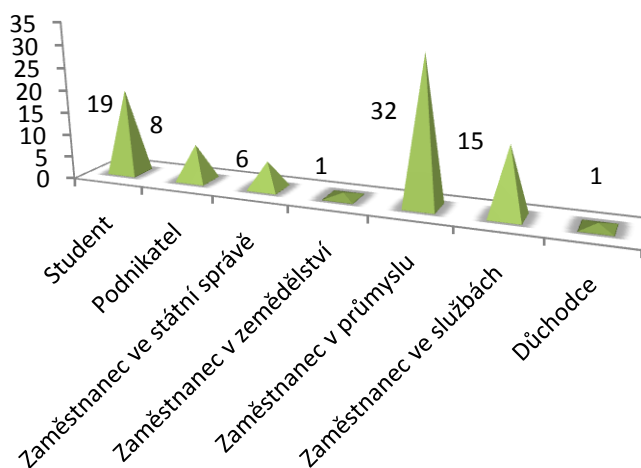


Zdroj: Vlastní tvorba

Tento graf ukazuje, že i respondenti, kteří využívají tohoto operátora, si ho pořídili pro osobní účely a jen opět velmi malý počet ho využívá i pro firemní účely.

f) Poslední čím byli dotazovaní ovlivněni, bylo jejich povolání.

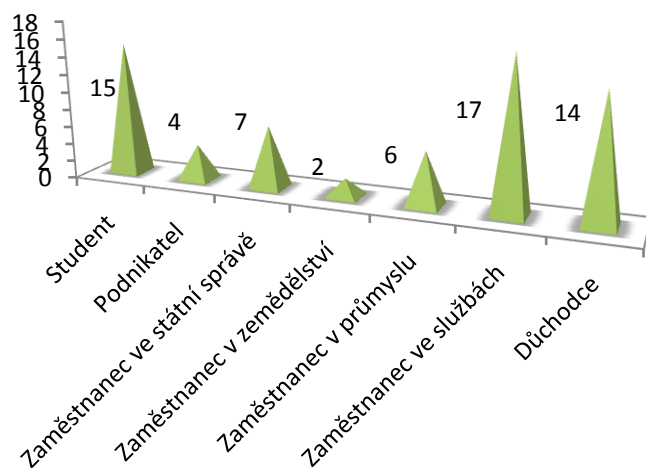
Graf 35: Povolání zákazníků Telefónici O2



Zdroj: Vlastní tvorba

Telefónicu O2 využívá nejvíce okruh respondentů zaměstnaných v průmyslu. Opět je to z toho důvodu, že s touto společností má uzavřenou smlouvu společnost ČEZ, která je průmyslovou firmou. Dále tohoto operátora využívá i větší počet studentů a zaměstnanců ve službách.

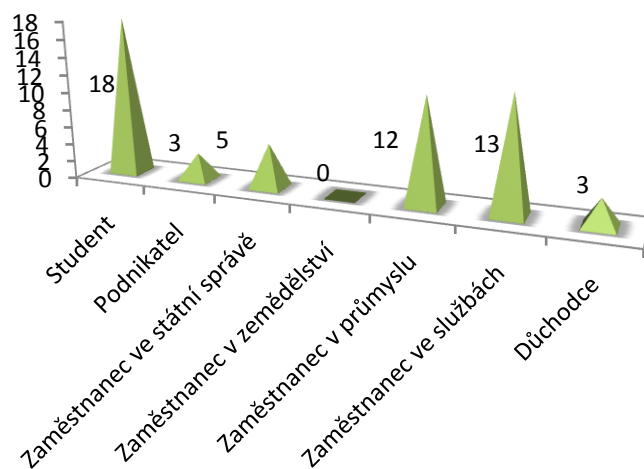
Graf 36: Povolání zákazníků T-Mobilu



Zdroj: Vlastní tvorba

T-Mobile nejčastěji vlastní skupina respondentů pracujících ve službách, dále potom studenti a také důchodci.

Graf 37: Povolání zákazníků Vodafonu



Zdroj: Vlastní tvorba

Vodafone je nejvíce oblíben mezi skupinou respondentů, kteří stále studují. Studenti mají omezené finanční možnosti a tento operátor nabízí své služby za nejnižší ceny. Dále ho také využívá větší počet zaměstnanců ve službách a také v průmyslu.

Předchozí tři grafy ukazují, že Telefónicu O2 nejvíce využívá okruh respondentů pracujících v průmyslové oblasti, T-Mobile okruh respondentů z oblasti služeb a Vodafonu dávají nejvíce přednost studenti pro jeho nejnižší ceny.

5 ZÁVĚR:

Ve své diplomové práci jsem se zabývala využitím klasifikačních stromů v marketingovém průzkumu trhu, přičemž jsem se zaměřila na oblast poskytování mobilních služeb.

Cílem mé práce bylo pomocí progresivní metodologie CART (klasifikační a regresní stromy) identifikovat důležité faktory ovlivňující chování zákazníka.

V období od 1. 12. 2007 do 31. 1. 2008 jsem prováděla dotazníkové šetření a na základě získaných dat jsem dospěla k závěrům, že zákazníky mobilních operátorů nejvíce ovlivňují tyto faktory:

- 1) Čistý měsíční příjem, kdy jsem dospěla k názoru, že zákazníci s vyššími příjmy preferují spíše mobilního operátora Telefónica O2 a naopak lidé s nižšími příjmy dávají přednost T-Mobilu a Vodafone.
- 2) Věková kategorie, kdy zákazníci z nižší věkové kategorie využívají především Vodafone a T-Mobile a zákazníci ze střední věkové kategorie potom preferují Telefónicu O2.
- 3) Respondenti se také rozhodují pro koupi mobilního operátora na základě toho, zda je pro ně prioritní cena, značka, doporučení či nabízené služby. Zákazníci se pro koupi jakéhokoli mobilního operátora rozhodují nejvíce podle ceny.
- 4) Služby, které zákazníci nejvíce využívají, kdy nejširší okruh respondentů odpověděl, že nejvíce využívá volání a posílání SMS zpráv.
- 5) Respondenty také ovlivnilo to, zda mobilního operátora používají pro osobní či firemní účely. Většina dotazovaných využívá mobilní operátory jen pro osobní účely. Největší počet zákazníků, kteří využívají mobilního operátora i pro firemní účely má Telefónica O2. U ostatních mobilních operátorů je tento počet zanedbatelný.

- 6) Povolání, kdy Telefónicu O2 nejvíce využívá okruh respondentů pracujících v průmyslové oblasti, T-Mobile okruh respondentů z oblasti služeb a Vodafonu dávají nejvíce přednost studenti pro jeho nejnižší ceny.

Mým doporučením pro další využití vyhodnoceného dotazníku je zaobírat se marketingovým mixem, nebo-li souborem taktických marketingových nástrojů - výrobní, cenové, distribuční a komunikační politiky, který společnosti umožňuje upravit nabídku podle přání zákazníků na cílovém trhu. Na základě vyhodnoceného dotazníku si může firma zjistit, kdo jsou její zákazníci a roztrždit je do segmentů. Součástí segmentace je i zjištění, na co jednotlivé skupiny zákazníků reagují a jaké nejvíce osloví nabídka dotyčné firmy. Dále může firma zjistit, jak se nejlépe dostat ke svým zákazníkům a jak je zaujmout a nakonec se může rozhodnout, jak se bude prezentovat, na co klást důraz a s jakou vlastností či vlastnostmi si ji budou zákazníci spojovat.

6 SUMMARY:

In my thesis I dealt with usage classification trees in market analysis, whereas I focused on area providing mobile services.

The aim of my work was, by use of progressive methodology CART (classification and regression trees), to identify important factors biasing consumer behavior.

On the basis of this questionnaire inquiry I have come the opinion, that consumers purchase mobile phone services base on their net monthly income, age, occupation, services required and whether they use the mobile service for personal or business purposes.

7 POUŽITÁ LITERATURA:

- [1] www.vodafone.cz
- [2] www.t-mobile.cz
- [3] www.cz.o2.com
- [4] William A. Belson: Matching and Prediction on the Principle of Biological Classification, Applied Stat., 1959.
- [5] James N. Morgan, John A. Sonquist: Problems in the Analysis of Survey Data, and a Proposal, Journal of the American Statistical Association, 1963.
- [6] Experiments in Induction – E. B. Hunt, J. Marinová a P. J. Stone, 1966
- [7] Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C. J.: Classification and Regression Trees. Wadsworth, 1984
- [8] Savický, P., Klaschka, J., Antoch, J.,: Optimální klasifikační stromy, Robust 2000
- [9] cmp.felk.cvut.cz/cmp/courses/recognition/
- [10] Klaschka, J.,: Lesk a bída optimálních stromů, Robust 2002
- [11] Feldesman, MR., Classification Trees as an Alternative to Linear Discriminant Analysis, American Journal of Physical Anthropology, 2002

8 SEZNAM OBRÁZKŮ, TABULEK A GRAFŮ:

8.1 Seznam obrázků:

Obrázek 1: Klasifikační strom str. 13

8.2 Seznam tabulek:

Tabulka 1: Parametr složitosti str. 34

Tabulka 2: Parametr složitosti str. 39

8.3 Seznam grafů:

Graf 1: Míra návratnosti str. 19

Graf 2: „Jakého mobilního operátora využíváte?“ str. 21

Graf 3: „Pro jaké účely?“ str. 22

Graf 4: „Co pro Vás bylo při výběru mobilního operátora rozhodující?“ str. 23

Graf 5: „Dáváte přednost“ str. 24

Graf 6: „Které mobilní služby vyžíváte?“ str. 25

Graf 7: „Jakým způsobem jste si mobilního operátora pořídili?“ str. 26

Graf 8: „Vaše měsíční útrata za mobilní služby činí“ str. 27

Graf 9: „Pohlaví“ str. 28

Graf 10: „Nejvyšší ukončené vzdělání“ str. 29

Graf 11: „Věková kategorie“ str. 30

Graf 12: „Povolání“ str. 31

Graf 13: „Příjmová kategorie“ str. 32

Graf 14: Kompletní klasifikační strom vygenerovaný v rpart	str. 33
Graf 15: 1 SE pravidlo a CP	str. 35
Graf 16: Ořezaný klasifikační strom vygenerovaný v rpart	str. 36
Graf 17: Klasifikační strom vygenerovaný v rpart bez faktoru X6	str. 38
Graf 18: 1SE pravidlo a CP	str. 40
Graf 19: Ořezaný klasifikační strom vygenerovaný v rpart bez faktoru X6	str. 41
Graf 20: Čistý měsíční příjem zákazníků Telefónici O2	str. 45
Graf 21: Čistý měsíční příjem zákazníků T- Mobilu	str. 46
Graf 22: Čistý měsíční příjem zákazníků Vodafoneu	str. 46
Graf 23: Věková kategorie zákazníků Telefónici O2	str. 47
Graf 24: Věková kategorie zákazníků T-Mobilu	str. 48
Graf 25: Věková kategorie zákazníků Vodafoneu	str. 48
Graf 26: Rozhodující faktory zákazníků Telefónici O2	str. 49
Graf 27: Rozhodující faktory zákazníků T-Mobilu	str. 50
Graf 28: Rozhodující faktory zákazníků Vodafoneu	str. 50
Graf 29: Využívané služby zákazníky Telefónici O2	str. 51
Graf 30: Využívané služby zákazníky T-Mobilu	str. 52
Graf 31: Využívané služby zákazníky Vodafoneu	str. 52
Graf 32: Využívání Telefonici O2 pro osobní či firemní účely	str. 53
Graf 33: Využívání T-Mobilu pro osobní či firemní účely	str. 54
Graf 34: Využívání Vodafoneu pro osobní či firemní účely	str. 55

Graf 35: Povolání zákazníků Telefónici O2 str. 55

Graf 36: Povolání zákazníků T-Mobilu str. 56

Graf 37: Povolání zákazníků Vodafonu str. 57

9 PŘÍLOHY:

Seznam příloh:

- Dotazník

DOTAZNÍK:

MOBILNÍ OPERÁTOŘI V ČR:

Vážené dámy, vážení pánové.

Dovoluji si Vás kontaktovat a požádat o pomoc při zjišťování oblíbenosti mobilních operátorů mezi spotřebiteli v České republice. Veškeré Vámi poskytnuté informace budou využity pouze za účelem vypracování diplomové práce.

Za spolupráci předem děkuji.

Se srdečným pozdravem

Bc. Kateřina Prokopová

1. Jakého mobilního operátora využíváte?

- ☐ Telefonica O2
- ☐ T-Mobile
- ☐ Vodafone

2. Pro jaké účely?

- ☐ Osobní
- ☐ Firemní

3. Co pro Vás bylo při výběru mobilního operátora rozhodující?

- ☐ Cena
- ☐ Značka
- ☐ Doporučení
- ☐ Nabízené služby
- ☐ Jiné

4. Dáváte přednost

- ☐ Měsíčním paušálům

- ☐ Předplaceným kartám

5. Které mobilní služby využíváte?

- ☐ Volání
- ☐ SMS
- ☐ MMS
- ☐ Záznamové služby
- ☐ Roaming a volání do zahraničí
- ☐ Mobilní bankovníctví a platby
- ☐ Internet
- ☐ Wap
- ☐ E-mail a fax

6. Jakým způsobem jste si mobilního operátora pořídili?

- ☐ Osobně
- ☐ Telefonicky
- ☐ Přes internet
- ☐ Jiné

7. Vaše měsíční útrata za mobilní služby činí

- ☐ Méně než 250 Kč
- ☐ 250 – 500 Kč
- ☐ 500 – 750 Kč
- ☐ 750 – 1000 Kč
- ☐ 1000 – 1500 Kč
- ☐ 1500 – 2000 Kč
- ☐ 2000 – 3000 Kč
- ☐ Více než 3000 Kč

8. Pohlaví

- ☐ Žena
- ☐ Muž

9. Nejvyšší ukončené vzdělání

- ☐ Základní
- ☐ SOČ bez maturity
- ☐ SŠ s maturitou
- ☐ Vyšší odborné
- ☐ Vysokoškolské

10. Věková kategorie

- ☐ Do 15 let
- ☐ 16 – 18 let
- ☐ 19 – 26 let
- ☐ 27 – 35 let
- ☐ 36 – 45 let
- ☐ 46 – 55 let
- ☐ 56 – 65 let
- ☐ 65 a více let

11. Povolání

- ☐ Student
- ☐ Podnikatel
- ☐ Zaměstnanec ve státní správě
- ☐ Zaměstnanec v zemědělství
- ☐ Zaměstnanec v průmyslu
- ☐ Zaměstnanec ve službách
- ☐ Důchodce

12. Příjmová kategorie (Váš čistý měsíční příjem)

- ☐ Bez příjmu
- ☐ Do 10 tisíc Kč
- ☐ 10 – 15 tisíc Kč
- ☐ 15 – 20 tisíc Kč
- ☐ 20 – 25 tisíc Kč
- ☐ 25 – 30 tisíc Kč
- ☐ 30 – 35 tisíc Kč
- ☐ 35 – 40 tisíc Kč
- ☐ 40 tisíc Kč a více