

Ekonomická
fakulta
Faculty
of Economics

Jihočeská univerzita
v Českých Budějovicích
University of South Bohemia
in České Budějovice

Jihočeská univerzita v Českých Budějovicích

Ekonomická fakulta

Katedra aplikované matematiky a informatiky

Téma bakalářské práce:

Získávání zákazníků v e-commerce: Využití metod data mining

Student: Pavel Matička

Vedoucí bakalářské práce: Doc. Ing. Ladislav Beránek CSc.

Studijní rok 2017

Čestné prohlášení

Prohlašuji, že svoji bakalářskou práci jsem vypracoval samostatně pouze s použitím pramenů a literatury uvedených v seznamu citované literatury. Prohlašuji, že v souladu s § 47 zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své bakalářské práce, a to - v nezkrácené podobě v úpravě vzniklé vypuštěním vyznačených částí archivovaných Ekonomickou fakultou - elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách, a to se zachováním mého autorského práva k odevzdanému textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

Datum:

Podpis studenta:

Poděkování

Největší projev díku bych chtěl vyjádřit svému vedoucímu bakalářské práce panu Doc. Ladislavu Beránkovi, kterému bych chtěl poděkovat za ochotu, vstřícnost, trpělivost a mnohé věcné podněty, které mi z velké části pomohly dokončit tuto bakalářskou práci. Dále bych chtěl poděkovat Janu Matičkovi, za pomoc projevenou při zajišťování databáze. Velký dík patří mé rodině, za veškerou podporu, které se mi v průběhu studia dostávalo a stále dostává.

Abstrakt (CZ)

Mnoho společností si uchovává data o svých zákaznících a následně se s nimi snaží pracovat. Hlavním cílem této práce je analýza databáze uživatelů, která byla poskytnuta od providera telekomunikačních služeb. Tato databáze obsahuje rozsáhlé množství dat o zákaznících společnosti a to například pohlaví, věk, geografický původ, pořízenou službu a další užitečné informace.

Na databázi jsou následně aplikované různé metody data miningu, které vedou k rozdělení databáze jako celku na potencionální ekonomické skupiny zákazníků. Závěr této práce obsahuje tabulky a grafy, které vzešly z analýzy databáze. Celkový výzkum může být použit jako pracovní materiál pro poskytovatele telekomunikačních služeb, který může tyto informace využít ke zvýšení celkového prodeje svých služeb. V neposlední řadě jsou všechna data z výzkumu aplikována do Geografického Informačního Systému, který graficky znázorní data na území České republiky.

Klíčová slova: E-commerce, data mining, zákazník, analýza

Abstract (EN)

Most companies keep records of their visitors and then they are trying to turn them into customers. The main goal of this thesis is an analysis of a user database from one of an internet provider. This database contains extensive amount of data about visitors of internet websites, such as website browsing behaviour, geographic location and other important information.

Different methods of data mining are applied on this database in order to divide potential customers into different economic groups. The conclusion of the work contains tables and charts from the databases research, which can be used as a workflow material for the internet provider. The provider can use this information to increase specific product sales. Thereafter, all data are inserted into the Geographic information system, which draws a final set of users on the map of the Czech Republic.

Key words: E-commerce, data mining, customers, analysis,

Obsah práce

1. Úvod	6
2. Cíl práce	8
3. Analýza E-commerce	9
3.1 E-commerce.....	10
3.2 B-2-B Business to business	11
3.3 B-2-B ve firmě UVXA.....	12
3.4 B-2-C Business to customer	12
3.5 B-2-C ve firmě UVXA.....	15
4. Data mining	16
4.1 Dobývání znalostí z databází	17
4.2 Strojové učení – Machine learning.....	19
4.3 Supervised learning	20
4.4 Data miningové metody	21
4.4.1 Rozhodovací stromy – decision tree	21
4.4.2 Asociační pravidla	26
4.4.3 Aglomerativní shlukování	31
5. Praktická část.....	33
5.1 Databáze zákazníků firmy UVXA.....	33
5.2 Grafická analýza okresů databáze firmy UVXA v prostředí ArcGIS	36
5.3 Grafická analýza krajů databáze firmy UVXA v prostředí ArcGIS	37
5.4 Analýza Královehradeckého kraje	40
5.4.1 Rozhodovací strom Královehradecký kraj	42
5.4.2 Asociační pravidla Královehradecký kraj	44
5.4.3 Získání zákazníků v Královehradeckém kraji.....	45
5.5 Analýza Karlovarského kraje.....	46
5.5.1 Rozhodovací strom Karlovarský kraj	48
5.5.2 Asociační pravidla pro Karlovarský kraj.....	51
5.5.3 Získání zákazníků v Královehradeckém kraji	52
6. Závěr	54
7. Seznam použitých zdrojů.....	55
8. Seznam obrázků	56
9. Seznam tabulek	57
10. Seznam příloh.....	58
11. Přílohy	60

1. Úvod

Neustálý rozvoj technologií, cenová dostupnost zařízení, velkoplošná rozšířenost internetu pro jeho uživatele a mnoho výrobců, kteří jsou ochotni nabízet své produkty prostřednictvím E-shopů - to jsou hlavní faktory, které daly možnost rychlému rozvoji e-commerce, která je nedílnou součástí dnešního obchodního světa a tvoří velmi podstatnou část celkového prodeje většiny subjektů, které působí na trhu výrobků a služeb jako prodávající. Bod zlomu nastal pro e-commerce na konci minulého století a to konkrétně v 90. letech. V této době se začala dostupnost internetu rapidně zvyšovat, což mělo za následek celosvětovou revoluci v oblasti businessu, komunikace a ve své podstatě internet vyvolal i ekonomickou změnu uvažování u drtivé většiny spotřebitelů. Zároveň započal pokles cen u osobních počítačů, které začaly být finančně dostupné i pro firmy a domácnosti. Díky této skutečnosti začal počet uživatelů internetu intenzivně růst. Dle posledních údajů internet každodenně využívá v České republice přes 72 % populace. V současnosti je Česká republika pokryta z více jak 90 % internetovým připojením, což nám dává možnost nakupovat prakticky z jakéhokoliv místa (Suchánek, 2012).

Většina prodejců, kteří působí na území České republiky, začala na nárůst poptávky po internetových obchodech reagovat a počet registrovaných e-shopů začal rychle stoupat. Podle posledních údajů se počet internetových obchodů od roku 2010 zvedl o bezmála 80%. V současné době tedy existuje na našem území přes 38 000 registrovaných internetových obchodů. Bez povšimnutí nemůžeme nechat ani obrat, kterého e-shopy dosahují. Jen v roce 2015 dosáhl celkový obrat 81 miliard korun.

Koupě zboží na internetu je velmi často spjata také s registrací. Při registraci zákazník poskytuje společnosti ve většině případů své národné údaje, které jsou následně zaváděny do databází, se kterými je následně možno, v rámci dodržení ochrany osobních údajů, nadále pracovat. Většina firem, které se snaží udržet si svou silnou pozici na trhu, pak dále tato data využívá k lepšímu zacílení prodeje svých produktů na konkrétní zákazníky či na selekci produktů do kamenných obchodů v určitých lokalitách. Celková predikce ideálního zacílení produktu na určitou skupinu zákazníků může vést ke zvýšení zisků prodejce a k celkovému upevnění pozice na trhu.

Následující obsah bakalářské práce bude rozdělen do dvou sekcí. První sekcí je teoretická část. V teoretické části budou uvedeny informace týkající se problematiky e-commerce jakožto celku a následně budou rozebírány některé z jejich druhů. V této práci jsem se ohledně problematiky e-commerce rozhodl zaměřit jen na dva druhy a to konkrétně na druh B2B a B2C, které jsou dále definovány na následujících stránkách. Výběr B2B a B2C je učiněn záměrně, jelikož jsou to hlavní způsoby, jak se námi zkoumaný poskytovatel telekomunikačních služeb v internetovém prostředí prezentuje a působí na své zákazníky. Databáze, která bude v této práci analyzována pomocí data miningu, byla poskytnuta jedním z předních českých poskytovatelů telekomunikačních služeb. Firma za poskytnutí informací požaduje anonymitu a z tohoto důvodu nemohu uvést, o jakou firmu se jedná. Pro lepší pochopení v textu bude firma nazývána zkratkou UVXA, která nijak nesouvisí s názvem společnosti. Dalším pokračováním teoretické části je popis fungování dobývání znalostí z databází a následně data mining, který je klíčový pro rozbor databáze. V neposlední řadě je teoretické sekci uveden popis jednotlivých metod, jejichž využití se objeví v praktické části.

Praktická část práce bude zaměřena na analýzu databáze. Databáze obsahuje nejnovější údaje o veškerých zákaznících firmy za rok 2016. Databáze je tvořena z dat, které pochází z údajů o pořízených službách jednotlivými spotřebiteli, nikoliv firmami. Tím pádem se jedná o prodej, který je zacílen čistě na jednotlivé koncové zákazníky. V praktické části jsou použity dva hlavní softwary pro analyzování databáze a to sice Geografický Informační Systém (GIS) a RapidMiner. GIS slouží v praktické části pro vizualizaci určitých prvků databáze a program RapidMiner je základem pro aplikování data miningových metod. V závěru praktické části je provedena analýza slabých krajů a následné obchodní doporučení firmě UVXA, založené na získaných informacích pomocí využití metod data miningu.

2. Cíl práce

Cílem bakalářské práce je analýza databáze uživatelů nejmenované společnosti, poskytující telekomunikační služby. Pomocí metody data mining a s využitím volně dostupného softwaru RapidMiner jsou predikovány jednotlivé skupiny potencionálních zákazníků, na které lze zacílit specifikovaný druh služby. Před začátkem rozboru databáze je nastíněna problematika E-commerce a veškeré data miningové techniky, které jsou v praktické části použity. Praktická část se zabývá nejdříve plošnou analýzou databáze, která určí nejméně ziskové kraje pro firmu UVXA. Ekonomicky nejslabší kraje jsou následně podrobně analyzovány metodou data miningu a na základě získaných informací je firmě UVXA doporučena ideální obchodní taktika ve slabých krajích.

3. Analýza E-commerce

Následující kapitola popisuje druhy e-commerce coby celek a také některé její druhy, v závislosti na tom, mezi jakými subjekty jednotlivé obchodní transakce probíhají. U každého obchodního modelu je uveden popis jeho fungování a následně je nastíněno, jakým způsobem firma daný obchodní model využívá. V závislosti na znalostech ohledně firmy UVXA jsem se rozhodl uvést hlavní dva modely internetového obchodování. Konkrétně je to model business to business, zkráceně označován také jako B2B a business to customer, pro který se používá zkratka B2C. Firma UVXA je na českém trhu velmi významným „hráčem“ v oblasti poskytování telekomunikačních služeb a k dosažení silné pozice ji velmi napomohlo provozování e-shopu. Na e-shopu firmy lze najít kompletní nabídku jejich služeb a případně provést objednávku služby/produktu. Celý proces objednávky se odehrává ve velmi přehledném uživatelském prostředí a zabere jen několik málo minut. Po zaslání požadavku na konkrétní druh služby/produktu je zákazník ještě tentýž den osobně kontaktován a jsou s ním upřesněny podmínky obchodu. Přístup UVXA ke svým zákazníkům tedy lze považovat za velice profesionální a dle mého názoru je to jeden z důvodů úspěšnosti firmy na českém trhu.

Díky informacím, které jsem získal při rozhovoru s členem marketingového oddělení firmy UVXA si trůfám tvrdit, že se firma orientuje spíše na koncového uživatele a tím pádem na obchodní model B2C. Samozřejmě má UVXA své zákazníky i v oblasti B2B, těch je však podstatně méně a nejsou pro ni hlavním příjmem finančních prostředků. Orientace na koncové zákazníky a tedy využívání modelu B2C je zřetelně vidět i na databázi zákazníků, jež mi byla od firmy UVXA poskytnuta. Obsahuje totiž přes půlmilion zákazníků, kteří mají zakoupenou minimálně jednu z celkově nabízených čtyřiceti služeb. V informacích o zákaznících firmy UVXA však převládá možnost, kdy si spotřebitel vybere více nabízených služeb najednou a dostane možnost aplikovat na celkovou sumu slevu. Tato nabídka je buďto řešena se zákazníkem individuálně, nebo si kombinaci služeb a přesnou výši slevy může zjistit na e-shopu, což výrazně urychluje proces poskytnutí objednávky a snižuje náklady firmy.

3.1 E-commerce

E-commerce lze jednodušeji popsat jako obchodování na internetu. Je to určitá forma obchodování, která k prodeji, nákupu a získávání znalostí o zákaznících využívá elektronickou síť. Používání této formy obchodování je pro prodejce i zákazníka oboustranně přínosné. Hlavním předpokladem pro fungování e-commerce je existence e-shopů a pokud možno i velké množství prodejců, kteří jsou ochotni pro zvýšení svých prodejů založit internetový obchod. Za tohoto stavu nabídky prodejců závisí pak jen na poptávce zákazníků, která je odvozena od cen, za které jsou schopni prodejci své zboží na internetu nabízet.

Možnost prodeje produktů na internetu má výhody pro obě strany, pro stranu prodávajícího tak i pro stranu kupujícího. Prodejce má díky e-shopu možnost snížit své náklady. Jedná se konkrétně o náklady, které se váží k provozování kamenného obchodu. „Klasická“ prodejna má hned několik nevýhod. Provozovatel musí vynaložit výdaje, které jsou s takovým provozováním spojené a to například platby za energie, nájemné, mzdu zaměstnanců atd. Provozování internetového prodeje tedy pro každého prodejce skýtá možnost ušetřit výdaje, což ale není, dle mého názoru, tím pravým kritériem. Hlavním kritériem je totiž možnost přilákat mnohem větší potencionální skupinu zákazníků, kteří se nemusí fyzicky pohybovat v kamenné prodejně, ale zároveň mohou nakupovat zboží, které prodejce na svých stránkách nabízí.

E-commerce však zahrnuje mnohem širší škálu možností využití než jen založení e-shopu. Mezi e-commerce lze zařadit i veškeré obchodní nabídky, se kterými se můžeme na internetu potkat. Do této skupiny pak tedy spadají i například internetové aukční portály, webové stránky nabízející k prodeji své služby, či bannery a online reklamy na zboží jednotlivých firem. Aktuální internetový trh se zbožím je velmi rozsáhlý a za takovéto situace přináší výhody pro zákazníky i prodejce. Hlavní výhodou pro zákazníka je nakupování z pohodlí domova a ve velkém případě nižší cena zboží na internetu, než v kamenných obchodech. Pro prodejce to pak znamená hlavní výhodou v ušetření nákladů a při případných registracích zákazníka také získání údajů, ze kterých se následně může snažit pomocí různých metod predikovat případné ideální nasazení produktů do kamenných prodejen v cílové lokalitě.

3.2 B-2-B Business to business

Business to business je proces obchodních transakcí, které probíhají mezi dvěma firmami či podniky. Ve své zásadě, je to jeden z nejstarších procesů elektronického podnikání a je založený na vzájemných interakcích dvou či více podnikatelských subjektů. Jednotlivé podnikatelské subjekty si z pravidla mezi sebou nabízí a vyměňují své produkty, služby a ostatní obchodní informace. Hlavním rozdílem mezi B2B modelem a B2C modelem je fakt, že v B2B modelu je za cílovou skupinu vždy označena firma, přičemž u B2C modelu je to vždy zákazník. Pro firmu nebo společnost, která vystupuje v obchodním vztahu jako poskytovatel produktů, či služeb, přináší obchodní model B2B mnoho výhod. Hlavní výhodou je povětšinou větší objem prodaného zboží/služeb při provedení jedné zakázky. Poptávající firmy jsou velice často označovány také jako velkoobchodníci či dealeři, kteří nakoupené zboží následně nabízejí k prodeji dále (retail). Retailový, neboli maloobchodní trh, se liší cenou, za které nakoupené produkty nabízí svým zákazníkům. Prodejní cena jednotlivých dealerů a velkoobchodníků je vždy navýšena o marži, ze které prodejce profituje a získává svůj zisk. Pro fungování celého systému business to business je velice důležité, aby jednotliví prodejci měli pokud možno co nejvíce údajů o velkoobchodnících, kterým své produkty nabízí (Suchánek, 2012). Proto je pravidlem, že každý zákazník, který chce odebrat zboží za velkoobchodní slevy, musí nejprve zažádat o registraci a následně ji dokončit například podpisem smlouvy, či online ověřením v případě registrace na webových stránkách prodejce. Druh této komerce je úzce spjat s možností náhledů spolupracujících firem do svých databází produktů. To by však nebylo možné, kdyby neexistovaly aplikace a software, který je schopen takovéto interakce mezi dvěma firmami zprostředkovat. B2B aplikace ve své podstatě lze připodobnit k běžným elektronickým obchodům. Hlavním rozdílem je fakt, že každá firma může mít stanovené určité podmínky, na základě kterých, provádí obchody v B2B aplikacích. Ve velké míře se již vyskytují i B2B aplikace, které jsou schopné provádět skladovou evidenci zásob, či automatické objednávky.

Poslední informace z roku 2014 mluví o 38 000 zaregistrovaných e-shopech na území České republiky (Tyden, 2016). Toto číslo však není úplně přesné. Je nutno poznamenat, že ne všechny e-shopy jsou po celou dobu aktivní, což se například týká transakčních webů. V dnešní době je samozřejmostí, že internetový obchod je schopný přímo reagovat s podnikovými informačními systémy. Tento fakt nahrává nejvíce právě firmám, které

disponují vysokým obratem zboží. Celý proces komunikace mezi oběma stranami obchodního styku je mnohem rychlejší, ale co víc, přináší zpětné vazby oběma stranám, a to v minimálním časovém intervalu. Internetový obchod lze napojit defacto na veškerý software, který se stará o plynulejší a časově méně náročný chod firmy. Velmi oblíbenou volbou spojení je například spolupráce se skladovými, účetními, ekonomickými a CRM systémy. Veškerá komunikace mezi těmito moduly pak probíhá povětšinou prostřednictvím webových služeb (Web services, API). Pokud se zaměříme na rozdíl v grafické podobě e-shopů u B2B a B2C je mezi nimi několik rozdílů, které plynou z požadavků firem, či zákazníků. Řešení, které se využívají pro B2B modely jsou tedy navrženy tak, aby bylo co možná nejjednodušší je spravovat a pružně reagovat na jakékoliv změny. Povětšinou jsou omezeny o reklamy a pro uživatele příjemné grafické prostředí.

3.3 B-2-B ve firmě UVXA

Poskytovatel telekomunikačních služeb, od kterého následně analyzovaná databáze pochází, působí v České republice již po mnoho let a v současné době zaujímá na telekomunikačním trhu jednu z hlavních pozic. Nabídka pro připojení ke svým službám je tato společnost schopna provádět na celém území České republiky, a to zejména díky dokonalé optické síti. Pokud se zaměříme na zkoumanou firmu a její fungování v business to business modelu zjistíme, že nabídka, kterou poskytují ostatním velkoodběratelům, je velmi široká. V oblasti poskytování telekomunikačních služeb je možno považovat využití modelu B2B za samozřejmost. Na našem území působí velké množství poskytovatelů, kteří mají své služby, jež následně prodávají koncovým zákazníkům, nakoupené od hlavních providerů. Telekomunikační společnost, která vystupuje v této bakalářské práci, stojí v B2B na straně prodejce, tudíž je jasné, že má na českém trhu značnou tržní sílu.

3.4 B-2-C Business to customer

Business to customer je model obchodních transakcí, které probíhají mezi prodejcem a konečným zákazníkem, přičemž hlavní iniciativa k prodeji služeb/produktů vychází od prodejců či poskytovatelů a cílovou skupinou jsou vždy spotřebitelé. Stejně jako u obchodního modelu B2B je i model B2C založen na prodeji, který probíhá na internetu. Nezbytným faktorem pro fungování obchodního modelu B2C je existence internetových obchodů (e-shopů). Prostřednictvím e-shopů se pak následně prodejce snaží, co možná

nejlépe, inzerovat své produkty/služby a nabízet je svým potencionálním zákazníkům, kteří e-shop navštíví. Ve své podstatě je veškerý software, který v dnešní době řeší fungování internetových obchodů, založen na velmi podobném principu jako software, který řeší fungování B2B. Vzhledem k tomu, že softwarová řešení pro business to customer začala vznikat ve chvíli, kdy byly internetové obchody pro B2B již velmi pokročile rozvinuty, získaly e-shopy pro B2C mnoho užitečných funkcionalit, které přebraly od svého předchůdce. Mezi naprosto samozřejmé funkcionality patří například konektivita e-shopu a podnikového informačního systému provozovatele.

Přímé napojení e-shopu na podnikový informační systém přináší mnoho výhod a úsporu času pro firmu. Velmi praktické je například propojení na účetní systém a evidenci skladů. Při každé nové objednávce jsou pak automaticky vystavovány faktury a zjišťován aktuální stav zboží na skladě. Automatizace procesů při vyřizování objednávek také velmi napomáhá k snížení chybovosti z hlediska selhání lidského faktoru. Takové chyby mohou nastat například když zaměstnanec zadává informace do účetního softwaru. Celková automatizace procesů je přínosná i z hlediska finanční stránky pro podnik, který může výrazně snížit své provozní náklady tím, že nemusí zaměstnávat pracovníky, za které vykonává funkci podnikový informační systém.

Hlavním rozdílem mezi internetovými obchody B2B a B2C je uživatelské prostředí a také celá grafická podoba. *„Cílem B2C e-shopů je získání a udržení zákazníka. Grafika obchodů pak musí oslovovat cílovou klientelu (ženy, muži, sportovci, důchodci, rybáři apod.), reprezentovat nabízené komodity zboží a celkové pojetí musí vytvářet příjemné prostředí pro nákup a snadnou orientaci zákazníka (ergonomie). Obvyklé jsou rovněž funkcionality (moduly), které mají za úkol co nejvíce prodávat způsobem, která zákazník neuráží, ale má vysoký výsledný efekt (akční zboží, výprodeje, novinky, související a alternativní zboží apod.)* (B2C, 2012). Nejde však jen o grafickou stránku. Procházení e-shopu musí být pro potencionálního návštěvníka jednoduché a přehledné. Stejně jako u B2B modelu, tak i v B2C je v naprosté většině případů požadována registrace zákazníka před tím, než je mu objednávka potvrzena a zboží odesláno ze skladu na jeho doručovací adresu. Řada e-shopů již dnes funguje i tak, že zboží zasílá bez registrace uživatele, to je však zákazník stejně nucen zadat doručovací údaje. Právě tato registrace či vyplňování

doručovacích údajů nesmí být zdlouhavé a náročné operace, jinak by potenciální zákazník mohl stránky prodejce opustit a vyhledat pro svůj nákup jinou společnost.

Jak již bylo zmíněno výše, v České republice existuje přes 38 000 zaregistrovaných e-shopů. Konkurence v odvětví obchodování na internetu je extrémně vysoká, a proto je nutné dbát nejen na správnou grafickou stránku a přívětivé prostředí pro návštěvníka, ale také na marketing. Pro zvýšení zisků z prodeje přes e-shop je využíváno mnoho nástrojů, které jsou integrovány přímo do nabízeného softwarového řešení internetového obchodu. Dle mého názoru, je nejpodstatnějším nástrojem SEO optimalizace. SEO optimalizace je označení pro metodiku, která se stará o takovou strukturu webových stránek, která je vhodná pro internetové vyhledávače. Pokud jsou webové stránky dobře navrženy a jejich struktura je optimální, internetové vyhledávače je upřednostní a tím pádem se zvyšuje možnost, že návštěvník zamíří právě na web, který je vhodně optimalizován (Suchánek, 2012). Dalším marketingovým nástrojem pro zvýšení efektivity prodeje je možnost zasílání e-mailů, ve kterých je povětšinou uvedeno akční zboží, či jiné aktuální benefity, které může návštěvník čerpat při nákupu na daném e-shopu. Zasílání propagačních e-mailů je povětšinou spojeno s registrací, kde má návštěvník na výběr, zda má o zasílání zájem či nikoliv. Provádění obchodních transakcí přes internet s sebou přináší mnoho výhod, a to jak pro koncového zákazníka, tak i pro podnik. Koncový zákazník ve velkém množství případů může profitovat z nižší ceny, než která by byla u stejného zboží nabízená v kamenném obchodě. Zároveň má díky e-shopům nespočet nabídek zboží/služeb a tím pádem má spotřebitel lepší možnost, jak maximalizovat svůj užitek z pořízení produktu. Pro podnik je provozování e-shopu záležitostí, kterou šetří svůj čas i peníze. Zboží, které je nabízeno v e-shopech je dostupné z jakéhokoliv místa, a tím pádem může podnik svou nabídkou produktů působit plošně na libovolnou množinu potenciálních zákazníků. Největší úspora pro podnik tedy plyne z faktu, že podnik nemusí otevírat kamenné obchody, a tudíž pádem platit poplatky spojené s provozováním prodejny, kterými jsou například mzdové výdaje či nájemné. Velkou výhodou pro podnik je možnost shromažďování informací o svých zákaznících. Veškeré údaje o zákaznících a provedených obchodních transakcích jsou zanášeny do databází, ve kterých jsou uchovávány a připraveny k následnému výzkumu.

3.5 B-2-C ve firmě UVXA

Obchodní model B2C je pro poskytovatele telekomunikačních služeb prioritním způsobem, kterým se snaží na internetu provádět své obchodní operace. Pokud zadáme do vyhledávače námi zkoumanou firmu zjistíme že se nachází hned na předních místech nejlepší shody. Tato skutečnost je daná zejména tím, že firma velmi pečlivě dbá na propagaci svých produktů a nabízených služeb. Přední místa samozřejmě obsazuje například díky ideální SEO optimalizaci a reklamě na vyhledávači Google, za kterou vynakládá peněžní prostředky. Při navštívení webových stránek provozovatele zaujme hned na první pohled příjemné grafické prostředí a celkové rozložení stránky. Hlavní navigační panel zabírá odkazy na e-shop, který nabízí různé produkty, které firma poskytuje. Každý produkt je podrobně popsán a doplněn o příjemnou grafickou podobu jednotlivých oken produktů. Objednávání služby/produktu je naprosto jednoduché a nezabere více jak 2 minuty. Při provádění objednávky máte možnost vytvořit si svůj uživatelský účet, pomocí kterého se nadále přihlašujete do databáze a tak můžete spravovat svůj účet. Plná platnost registrace však proběhne až ve chvíli, kdy má zákazník podepsanou smlouvu s poskytovatelem. Následně zadá své příjmení a číslo smlouvy a od té doby může spravovat svůj účet a objednávat další služby výhradně pomocí internetového e-shopu poskytovatele telekomunikačních služeb. V závislosti na výše zmíněných faktech, firma získává díky těmto registracím velké množství informací, které jsou ukládány do společné databáze. Ta je doplněna pracovníky firmy i o zákazníky, kteří podepíší smlouvu, ale nezaregistrují se na e-shopu. Souhrnně řečeno tato firma má veškeré záznamy o svých zákaznících a jejich službách, které poptávají. Této skutečnosti velmi napomáhá obchodní model B2C a také existence propracovaného a sofistikovaného internetového obchodu.

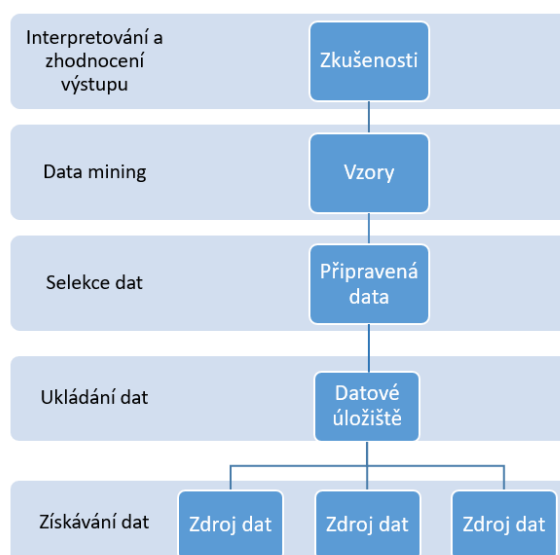
Databáze uživatelů, která bude zkoumána v praktické části této bakalářské práce je zaměřena právě na jednotlivé cílové zákazníky, nikoliv na firmy, jak je to u modelu B2B. Jelikož poskytovatel telekomunikačních služeb má na českém trhu velmi početnou skupinu zákazníků, jsou pro něj veškerá získaná data velice cenná a díky jejich shromažďování na těchto datech může nadále provádět výzkum. Správně provedená analýza databáze může poskytovateli telekomunikačních služeb přinést užitečné informace, pomocí kterých může například zvýšit svůj celkový zisk.

4. Data mining

Potřeba zkoumání závislostí mezi daty každým rokem stoupá. V dnešní době jsme prakticky zahlceni množstvím dat, které se lze získávat nespočetně způsoby a prakticky ve všech pracovních odvětvích. Historie data miningu sahá do 90. let minulého století, kdy se objevily první zmínky o této metodě a to sice na konferencích o umělé inteligenci IJAI'89 a konferenci americké asociace umělé inteligence AAAI'91 a AAA'93 (Berka,2005). Celkový vývoj umělé inteligence a rychlý rozvoj cenově dostupné techniky zapříčinil enormní nárůst dat a kvůli tomu i procesu dobývání znalostí z databází. V ohledu uchovávání dat, která jsou pro data mining klíčová, nastal hlavní zvrat s příchodem cenově dostupných velkokapacitních disků a výkonných osobních počítačů. V tuto chvíli se uživatelům, kteří o informace z dat měli zájem, otevřely nové možnosti. V dnešní době může proces dobývání znalostí z databází provádět jakýkoliv uživatel, který má potřebnou techniku a znalosti. S rostoucím množstvím dat ale také výrazně klesá jejich pochopení lidmi, přičemž lidstvo hledá různé vzory dat již od svého vzniku. Lovci se například snažili hledat vzory v pomyslných datech, které jim přinášelo migrační chování zvířat a zemědělci hledali a stále hledají vzory v půdách, plodinách a počasí. A konečně, v případě podniků je jejich úkolem, aby byli schopni identifikovat příležitost ze vzorů chování zákazníků, které jim poskytují data o prodeji produktů/služeb. V případě této bakalářské práce a následného dobývání znalostí z databáze od poskytovatele telekomunikačních služeb, jsou okolnosti úplně stejné. Masivní nárůst e-shopů a následná možnost propojení s podnikovými informačními systémy. Tato fakta zapříčinila skutečnost, že data lze získat od každého zákazníka, který na daném internetovém obchodě provede objednávku. Data mining je ve své podstatě jenom jedna z částí při celém procesu dobývání znalostí z databází, jehož průběh bude na následujících stránkách přesněji popsán. Proces data miningu je aplikován až v předposlední části a slouží nám k podrobnému rozboru dat a k hledání společných zákonitostí. Díky rozboru dat pak máme možnost nahlížet na celkovou databázi jako na jednotlivé celky, které nám povětšinou přináší mnoho různých informací.

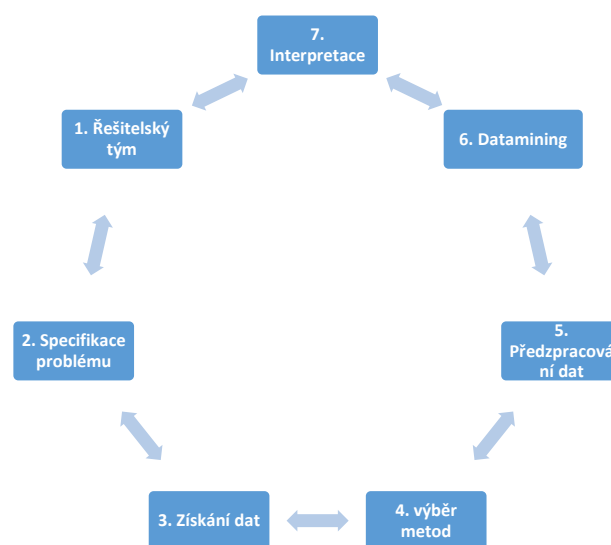
4.1 Dobývání znalostí z databází

Dobývání znalostí z databází, také (Knowledge Discovery in Databases, KDD), bylo definováno jako proces netriviálního objevování implicitních, dopředu neznámých a potenciačně použitelných znalostí v datech (Bramer, 2007). Data mining je pouze jednou částí z celého procesu dobývání znalostí z databází. Pokud se zaměříme na celý proces z technologické stránky věci, lze jej rozdělit do několika hlavních bodů. Prvním bodem je snaha o získání dat, která jsou následně zkoumána. Všechna příchozí data jsou integrována a následně jsou uložena do datového skladu neboli na datové úložiště. V dalším kroku probíhá selekce druhu dat, která jsou pro následné analyzování požadována. Tento proces probíhá zejména kvůli tomu, že v datovém skladu se mohou nacházet informace z různých odvětví a také v jiných formátech. V případě nesourodosti dat by následující proces data miningu byl znehodnocen, a proto je selekce dat důležitým článkem postupu. Výstupem ze selekce jsou připravená data, která jsou povětšinou ucelená v databázi, se kterou už je možné dále pracovat a aplikovat na ní různé metody data miningu. Výstupem celého procesu dobývání znalostí z databází jsou jednotlivé vzory, v našem případě jsou to například jednotlivé potenciační skupiny zákazníků. Jednou z nejdůležitějších částí celého procesu je závěrečné zhodnocení a interpretace výstupů, díky kterým získáváme nové znalosti. (Bramer, 2007).



Obrázek 1. Technologický postup výše uvedeného dobývání znalostí z databází (Bramer, 2007)

Na proces dobývání znalostí z databází se lze také zaměřit z manažerského hlediska. Jak uvádí paní docentka Mrázová, impulzem pro zahájení procesu je reálný problém (Mrázová, 2014), kterým je v případě této bakalářské práce predikce ekonomických skupin zákazníků poskytovatele telekomunikačních služeb. První zásadou je vytvoření řešitelského týmu, který tvoří experti na danou problematiku a to především v oblasti dat a jednotlivých metod dobývání znalostí z databází. Následně probíhá specifikace problému, ve kterém se skupina snaží rozdělit data do určitých skupin, popsat vzájemné zákonitosti a vytvořit tezi řešení. Další postup je už velmi podobný jako u technologického postupu. Řešitelský tým poté získává data, což může vést v některých případech, například při nedostatku informací v datovém úložišti a nesourodosti dat, i ke změně uvažování nad celým problémem (Berka, 2005). Dále probíhá výběr metody, která bude uplatněna při dobývání znalosti z databáze. Mezi používané metody například patří rozhodovací stromy, neuronové sítě, asociační pravidla a mnoho dalších. Po výběru metody se tým zaměří na předzpracování neboli selekci dat, která jsou převedena do standartního formátu, který daná metoda využívá. Na připravená data jsou aplikovány metody, které slouží k zjištění užitečných vztahů mezi daty, přičemž aplikování stejných či jiných metod může probíhat v libovolném počtu opakování. Konečnou fází je interpretace výstupů, které byly získány v průběhu celého procesu. Povětšinou se jedná o velké množství výsledků, které můžou být pro tým předvídatelné či naopak neočekávané a překvapivé. Veškeré výsledky je třeba přehledně uspořádat a prezentovat je v ucelené a logické formě (Mrázová, 2014).



Obrázek 2. Manažerský postup výše uvedeného dobývání znalostí z databází (Berka, 2005)

4.2 Strojové učení – Machine learning

Strojové učení je v dnešní době využíváno v mnoha odvětvích, u kterých je důležitá práce s velkými objemy dat, které například na první pohled nejeví jasné známky vzájemných zákonitostí. Lze jej označit za podoblast umělé inteligence, zabývající se algoritmy, umožňující umělé inteligenci učit se z dat a v určitých mezích se automaticky přizpůsobovat změnám. Využívání principů strojového učení je dnes již neodmyslitelně zapsáno do podvědomí společnosti. Praktické případy, které popisují využití strojového učení, jsou uvedeny níže v této kapitole. Aplikování strojového učení zůstávalo po dlouhou dobu výsadou korporací a výzkumných ústavů. Rychlý rozvoj průmyslu, elektroniky a obchodování na internetu však umožnilo poměrně rychlé získávání velkého množství dat i pro firmy. Tato skutečnost radikálně změnila četnost využívání strojového učení a jeho výhody začalo rychle využívat širší spektrum subjektů.

Použití strojového učení je tedy podmíněno větším počtem dat, než které by byl schopen zanalyzovat jedinec. Dalším požadavkem pro správné fungování je určení a definování cíle, kterého chce firma pomocí výzkumu dosáhnout. Hlavní výhodou, která z celého procesu je fakt, že firma využívající techniky strojového učení může získat výrazný náskok před svou konkurencí a tím pádem upevnit svou pozici na trhu. Průběh strojového učení probíhá následujícím způsobem. Veškerá data jsou rozdělena do dvou samostatných sekcí. První sekci lze nazvat trénovací sekci, která obsahuje tréninková data. Tréninková data se ve velké části případů skládají z párů vstupních objektů a očekávaného výstupu. V této části procesu se software učí a vytváří si z trénovacích dat model. Druhou částí dat jsou data testovací, na které je následně aplikován model z trénovacích dat a následně je ověřována funkčnost předem vytvořeného modelu.

Strojové učení lze rozdělit do několika různých kategorií v závislosti na tom, jakým algoritmem se řídí. Každý z algoritmů může využívat na vstupu stejná data za předpokladu, že je instance vhodná pro oba typy algoritmů. Hlavní myšlenkou, která zapříčiní rozdílnost obou metod strojového učení, je předpokládaný výstup ze zkoumání dat. Pokud je pro zadanou množinu vstupních dat předem zadán požadovaný výstup, pak se jedná o učení s učitelem, neboli supervised learning. Tento způsob učení má povětšinou dva způsoby výstupu a to v případě určené třídy klasifikaci a v případě určené hodnoty regresi. Druhým způsobem je učení bez učitele, neboli unsupervised learning. Tato metoda učení funguje na

principu neznámého výstupního formátu a hlavním požadavkem je, aby nám přinesla co možná nejširší škálu informací, které je možno z dat získat (Brammer, 2007). Poslední možností, která se k problematice strojového učení a používaných algoritmů váže je kombinace obou algoritmů. V této práci je hlavním využívaným algoritmem učení s učitelem, a proto na následujících stranách bude kladen větší důraz na jeho rozbor a následně na jeho použití v praktické části, která se zabývá analyzováním databáze firmy UVXA. Jak již bylo zmíněno výše, strojové učení je dnes již velmi úzce spjato s každodenním životem. Existuje mnoho případů, kde se techniky strojového učení uplatňují, přičemž jeho užitek pro návštěvníka webových stránek může mít různé formy. Strojové učení je hlavní využívanou technikou například na jakémkoliv typu emailových řešení, kde zastává formu rozlišování mezi nevyžádanou a vyžádanou poštou (Alpaydin, 2010) Další využití nalezneme například na internetových portálech, které nabízejí k prodeji různé zboží. Strojové učení zde vykonává činnost, díky které se potencionálnímu zákazníkovi objevují různé souvislosti ohledně prohlíženého a dříve zakoupeného zboží. Stejně tak je využíváno i v internetových vyhledávacích či na sociálních sítích. Pokrok umělé inteligence, a tím pádem i strojového učení je velmi rychlý a z tohoto důvodu můžeme její formy najít například i v softwarech pro rozpoznávání řeči nebo při predikci předvolebních průzkumů.

4.3 Supervised learning

Učení s učitelem je algoritmem strojové učení, při kterém probíhá analýza dat na základě odvozování funkcí z předem označených tréninkových dat (Alpaydin, 2010). Algoritmus, který řeší danou problematiku, funguje na principu párování vstupních dat a požadované výstupní hodnoty, často také označované jako signál dohledu (supervisory signal). Algoritmus, řešící úlohy učení s učitelem, analyzuje na základě párů dat odvozenou funkci a nadále jí používá k mapování nových příkladů (Alpaydin, 2010). V optimálním případě fungování algoritmu jsou nalezeny a určeny označené třídy. Tyto třídy jsou následně aplikovány na instance, které doposud nebyly zkoumány. Hlavním faktorem pro úspěšné fungování celého procesu je množství trénovací množiny dat. Obecně platí, že čím větší množina tréninkových dat je použita, tím si zvyšuje pravděpodobnost, že náhodně vybraná konzistentní funkce bude mít po aplikaci na jiná data odpovídající výstup. Ve své podstatě strojové učení a jeho výstupy fungují na dvou základních principech a to sice je schopné předpovídat určité číslo, nebo třídít data do různých kategorií. Každý z těchto principů

využívá jiné algoritmy, kterými k výsledkům dojde a některé z nich budou na následujících stranách objasněny.

Prvním případem, kdy výstupem z funkce je hodnota, povětšinou spojitá, je nazýván regresí. Druhý případ může predikovat označení třídy u vstupních dat. Tento druh výstupu se nazývá klasifikací a v sekci učení s učitelem je nejhojněji zastoupen. Podstatným faktem je, že učení s učitelem je nejvýznamnějším využívaným druhem strojového učení a některé z jeho metod neodmyslitelně patří k pojmu supervidovaného učení. Hlavními metodami, které budou v této práci vysvětleny a nadále používány jsou rozhodovací stromy a neuronové sítě.

4.4 Data miningové metody

4.4.1 Rozhodovací stromy – decision tree

Rozhodovací stromy jsou nejčastěji používané při řešení problémů, týkajících se strojového učení, dobývání znalostí z databází a potažmo i umělé inteligence. Dokáží řešit problémy spojené s každodenním životem i složitější typy problému, které jsou v našem případě zastoupeny predikcí zákazníků ve firmách. Rozhodovací stromy jsou velmi praktickým nástrojem strojového učení a to především pro úlohy, které se týkají klasifikace a predikce. Klasifikační úlohy fungují na principu označení cílového atributu v databázi, jež následně funguje jako hlavním kritériem pro učení a následné určování zákonitostí a vztahů mezi daty. Podoba cílového atributu může mít různé podoby, povětšinou je ale zadán v jednoduché binární podobě 0/1 – Ne/Ano. Tato proměnná může obsahovat například informaci, zda zákazník při návštěvě e-shopu provedl objednávku na produkt či nikoliv. Po spuštění procesu rozhodovacího stromu začne algoritmus vyhledávat mezi ostatními daty jednotlivé spojitosti, které vedou k finálnímu výsledku neboli k cílovému atributu. Rozhodovací strom, obohacený o informace z předešlého testování, následně můžeme aplikovat na jiné datové sady, u kterých například ještě stav cílového atributu neznáme. V závislosti na informacích, získaných z tohoto procesu, můžeme pomocí různých marketingových metod docílit zvýšení prodeje produktů.

Otázkou zůstává, proč je využití rozhodovacích stromů tak rozšířené a jak náročné je jej provozovat. Ve své podstatě jsou rozhodovací stromy vhodné i pro začátečníky, kteří se rozhodli pronikat do tajů strojového učení. Hlavní výhodou je fakt, že není úplně nutné, aby měl uživatel speciálně připravená data pro výzkum. Speciální přípravou dat je myšleno

například rozložení a převod dat na vektory a jiné způsoby úprav. Rozhodovací stromy jsou také schopné pracovat s jakýmkoliv druhy dat. Tím je myšleno, že dokáží zpracovávat jak číselné proměnné, tak i kategorické proměnné. Stromy jsou schopné si i samy určit, které z dostupných atributů jsou pro výzkum potřebné. Tato možnost není samozřejmostí u všech druhů algoritmů, jelikož u ostatních typů algoritmů je předem nutné důležité atributy vybírat. Automatická selekce atributů je velmi významná zejména v okamžiku, kdy máme pro jeden záznam více atributů. Databáze firmy UVXA, která je v této práci podrobena data miningové metodě, obsahuje ke každému záznamu přes čtyřicet záznamů a vlastnost selekce tak přináší poměrně hojnou úsporu času. Největší výhodou skrývají rozhodovací stromy ve formě znázornění svého výstupu. Jak název napovídá, výstupem z algoritmického procesu je grafické znázornění stromu, obsahující veškeré potřebné informace. Strom lze snadno interpretovat a na první pohled v něm najít potřebné zákonitosti a informace. Velká část ostatních algoritmů však nenabízí takto pohodlné řešení výstupu. V mnoha případech algoritmus pracuje na principu zvaném black-box, což znamená, že uživatel ve své podstatě často neví, jak model dospěl ke konečnému řešení (Berka, 2005).

Pro využití rozhodovacích stromů existuje hned několik softwarů, které přináší možnost aplikace metod rozhodovacích stromů na databáze. Veškeré dostupné programy lze rozdělit do dvou hlavních kategorií a to sice na programy sloužící ke komerčnímu použití a naopak k nekomerčnímu použití. Mezi nejuznávanější freewarové programy pro používání rozhodovacích stromů, a potažmo i data miningu, patří například WEKA, YALE a RapidMiner. Poslední zmíněný software je v této práci využíván i přes některé okolnosti, které brání jeho plnému využití. Konkrétním problémem je omezení velikostí vstupu při provádění procesů. Maximálně velikost databáze, která může být načtena při zavádění vstupu, je zde omezena na 10 000 řádků, což v případě analyzování databáze přináší mnoho zásadních problémů. V praktické části je kladen důraz na jednotlivé kraje České republiky a na hledání jednotlivých možností, které by mohly přilákat k firmě UVXA nové zákazníky. Počty zákazníků firmy však skoro u všech krajů vysoko přesahuje hranici 10 000 a z tohoto důvodu je vždy analyzovaná databáze omezena na maximální limitní počet řádků.

Ve využívaném softwaru RapidMiner je k nabídce hned několik druhů rozhodovacích stromů, které je možné aplikovat při rozboru databáze. Základní možností výběru je klasický rozhodovací strom, založený na principu predikce atributu označeného jako label, využívající

operátorů pro rozhodovací stromy. Velikost a větvení rozhodovacího stromu je úzce spjata s množstvím atributů, vybraných při importování datové instance. Každý jednotlivý atribut je následně v konečném grafickém znázornění prezentován jedním konkrétním vnitřním uzlem stromu.

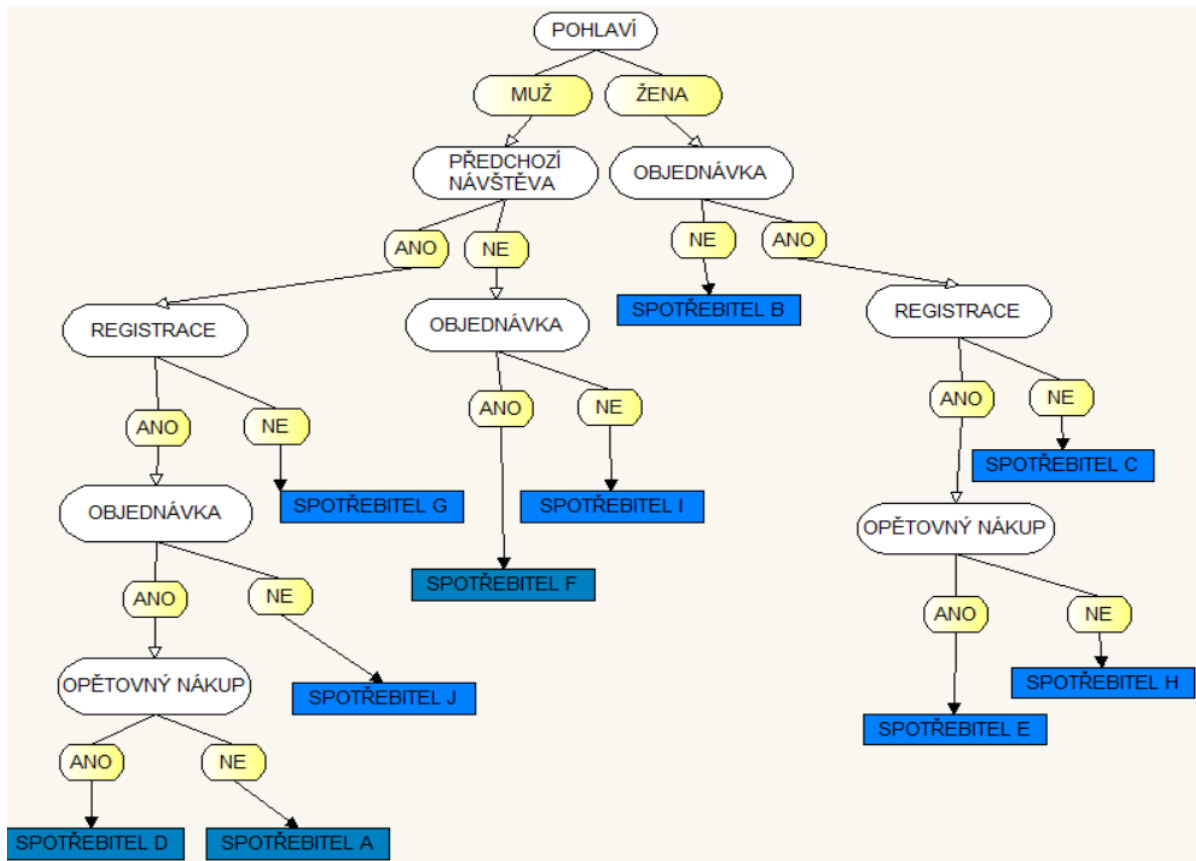
Z každého uzlu stromu vychází určitý počet hran, který je ovlivněn binárním zápisem hodnot, přidružených k danému atributu. Nejpoužívanějším druhem rozhodovacího stromu je v této práci ID3, neboli (Iterative Dichotomiser 3). ID3 funguje na principu využívání algoritmu generování rozhodovacího stromu, který vyvinul Ross Quinlan (Brammer, 2007). Fungování tohoto algoritmu je závislé na fixním vstupu datové sady, na základě které je rozhodovací strom budován. Každý list a uzel rozhodovacího stromu obsahuje označení (z pravidla binární 1/0 ANO/NE), zatímco koncový uzel přináší informace o konečné zkoumané veličině. Hlavním podstatným rozdílem u ID3 je využívání metod heuristiky pro ideální výběr atributů, jež jsou následně zařazeny do jednotlivých uzlů rozhodovacího stromu. Většina ostatních pravidel spojená se správným fungováním ID3 je velmi podobná již zmíněným podmínkám pro fungování klasického rozhodovacího stromu.

Zákazník	Předchozí návštěva	Registrace	Pohlaví	Objednávka	Opětovný nákup
Spotřebitel A	ano	ano	Muž	ano	ne
Spotřebitel B	ne	ano	Žena	ne	ne
Spotřebitel C	ne	ne	Žena	ano	ano
Spotřebitel D	ano	ano	Muž	ano	ano
Spotřebitel E	ano	ano	Žena	ano	ano
Spotřebitel F	ne	ano	Muž	ano	ano
Spotřebitel G	ano	ne	Muž	ne	ne
Spotřebitel H	ano	ano	Žena	ano	ne
Spotřebitel I	ne	ano	Muž	ne	ne
Spotřebitel J	ano	ano	Muž	ne	ne

Tab. 1 – Zadaná data pro ukázky rozhodovacího stromu (zdroj: autor)

Výše uvedená tabulka slouží jako malá datová instance pro následující ukázkou prvních dvou rozhodovacích stromů. Prvním případem je rozhodovací strom, vytvořený čistě na základě mého logického uvažování. Druhý případ je výstupem z prostředí RapidMiner, do kterého byla vložena k analyzování stejná databáze.

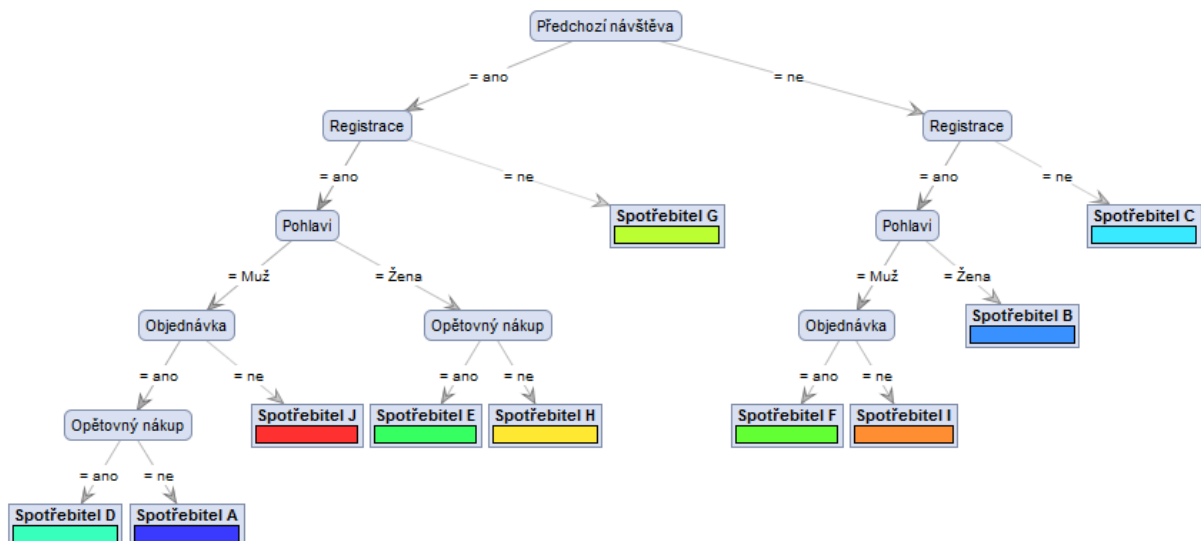
U obou případů byl jako label zvolen stejný atribut a to sice atribut zákazník. Struktura databáze a jednotlivé atributy by, dle mého názoru, mohly být velmi podobné například u jakéhokoliv druhu e-shopu. Na základě dostatečného množství dat a správného rozhodovacího stromu, může provozovatel e-shopu například určit chování svých zákazníků a případně zlepšit své nabízené služby, což může vést ke zvýšení jeho celkových příjmů.



Obr. 3 Grafické znázornění vlastního rozhodovacího stromu podle kritérií v tabulce 1 (zdroj: autor)

Na obrázku číslo 3 se nachází rozhodovací strom vytvořený čistě na základě subjektivního pocitu o důležitosti jednotlivých atributů. Jako prvním rozhodovacím atributem je zde označeno pohlaví uživatele, dělíci zákazníky do dvou jasných skupin. Stejně jako u všech následujících případů lze dvojici MUŽ/ŽENA zapsat binární kombinací 0/1. Použité označení je pouze z důvodu lepší přehlednosti. Následující popis je rozdělen do dvou odlišných sekcí, kde bude vysvětleno, z jakého důvodu byly různé atributy použity. Předpokládejme, že tabulka 1 představuje reálnou datovou sadu z e-shopu. Provozovatel e-shopu není spokojen se svými zisky, a proto se rozhodne zjistit jednotlivé chování zákazníků.

Bílé buňky v našem případě reprezentují jednotlivé atributy, žluté buňky určují hodnotu každé hrany a modré označení je použito pro konečnou množinu uživatelů – **label**. U muže nás jako první atribut zajímá hodnota u předchozí návštěvy webu. Kvůli hodnotám atributu jsou muži ihned rozděleni do dvou skupin, které se následně dělí dále. Pokud je zákazník na webu poprvé, zjišťujeme, zda při návštěvě e-shopu provede registraci či nikoliv. V tomto bodu rozhodovací strom zachybuje prvního zákazníka, který jako jediný neprovedl z mužů registraci. Při provedení registrace nás zajímá objednání zboží, u kterého je v dané situaci pouze jedna volba a to ano. Pro spotřebitele D a A je tedy limitujícím atributem opětovný nákup, na základě kterého jsou spotřebitelé rozděleni. U žen je proces velmi podobný, akorát je použita jiná hierarchie atributů. Rozdílnost atributů může být využívána v případě, že tvůrce rozhodovacího stromu má určité preference, které se snaží, vzhledem ke zvýšení efektivity prodeje, uplatnit.



Obr. 4 Grafické znázornění rozhodovacího stromu v programu RapidMiner (zdroj: autor)

Obrázek číslo 4 je na první pohled diametrálně odlišný od předchozího grafického náhledu rozhodovacího stromu. Společným pojítkem obou dvou znázornění jsou stejná data a ve své podstatě i označení atributu label. Důvodem jejich rozdílnosti je působení jednoho hlavního faktoru, a to sice již výše zmiňovaná vlastnost ID3, kterou je využití heuristiky. Algoritmus si tedy sám zvolil způsob klasifikace a určení potřebných atributů k dosažení výsledku. Při bližším zkoumání je však možné určitě zákonitosti v obou znázorněních najít, a to i za předpokladu, že rozhodovací strom na obrázku číslo 4 je vytvořen profesionálním

algoritmem a strom na obrázku číslo 3 je vytvořen pouze lidským úsudkem. Ve své podstatě algoritmus rozhodovacího stromu ID3 souhlasí prakticky se všemi koncovými atributy, které vedou k určení daného spotřebitele. Jediným rozdílem je případ spotřebitele B, kdy je atribut u určení spotřebitele jiný a to konkrétně „Pohlaví“ u obrázku číslo 4 a „Objednávka“ u obrázku číslo 3. Faktem však zůstává, že použitá ukázková množina dat obsahovala velmi málo dat, což umožnilo vytvořit rozhodovací strom na základě logického uvážení. Ve většině případů mají databáze enormní množství dat a je nutné je analyzovat pouze prostřednictvím příslušných počítačových programů. Pro upřesnění, ukázková sada dat obsahuje 66 buněk s hodnotami, oproti tomu databáze firmy UVXA obsahuje přes 19 000 000 buněk.

4.4.2 Asociační pravidla

Asociační pravidla jsou po rozhodovacích stromech dalšími prostředky, pomocí kterých je možné analyzovat různé zákonitosti v databázích. Nejčastější využití asociačních pravidel nalezneme zejména v analýze nákupního košíku, jež nám přináší vzájemné vztahy mezi produkty a četností jejich nákupu. Jedním z nejznámějších algoritmů asociačních pravidel je konstrukce IF-THEN, kterou je možné najít prakticky v každém programovacím jazyce (Berka, 2005). Při aplikaci asociačních pravidel na datové instance je povětšinou velmi důležitá správná selekce atributů a jejich rozdělení do podskupin. Podskupinou je v tomto případě myšleno rozdělení dat na antecedent a sukcedent. Pojem antecedent představuje předpoklad naší konstrukce IF-THEN a pojem sukcedent její závěr. Výhodou programu RapidMiner je fakt, že není nutné třídít atributy do jednotlivých podskupin, ale stačí je pouze vybrat. RapidMiner pak začne sám prohledávat veškeré kombinace podmínek pro atributy, na jejichž základě určí důležité zákonitosti mezi daty. Asociační pravidla jsou v této práci aplikovány jen velmi zřídka, a to z důvodu převážného využití rozhodovacích stromů, jejichž využití přináší podstatně lépe interpretovatelné výsledky. Nejpodstatnější využití asociačních pravidel je v této práci zejména na konce praktické části, kde nám jejich použití přináší komplexní informace o preferencích zákazníků firmy UVXA. Na základě získaných informací o preferencích je následně vytvořena teorie, jak by firma UVXA mohla získat nové zákazníky v problémových krajích.

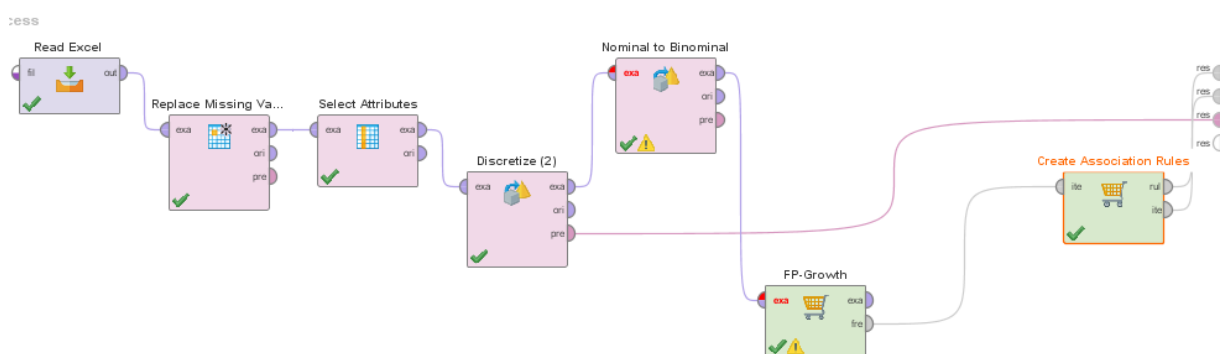
Pro lepší popis fungování asociačních pravidel je na následující straně znázorněn postup jejich fungování a také praktická ukáзка z prostředí RapidMiner. Použitá datová sada slouží stejně jako v kapitole o rozhodovacích stromech, pouze pro prezentaci fungování asociačních

pravidel a také pro popsání zákonitostí ve výstupní tabulce. Reálné případy použití na celistvou databázi jsou na následujících stranách této práce.

Zákazník	Analogová televize	Digitální televize	Internetové připojení	Hlasová služba
Spotřebitel A	1	1	1	0
Spotřebitel B	0	1	0	0
Spotřebitel C	0	1	0	0
Spotřebitel D	0	1	0	0
Spotřebitel E	1	1	1	0
Spotřebitel F	0	1	1	0
Spotřebitel G	0	1	0	0
Spotřebitel H	1	1	1	0
Spotřebitel I	1	1	1	0
Spotřebitel J	1	1	1	0

Tab. 2 – Zadaná data pro ukázkou fungování asociačních pravidel (zdroj: autor)

Tato tabulka obsahuje údaje o potencionálních zákaznících a kombinaci služeb, které si zakoupili. Použité služby (analogová televize, digitální televize, internetové připojení a hlasová služba), plně korespondují s pilotními službami nabízené firmou UVXA. Každý jednotlivý atribut služby má několik proměnných, které jsou uvedeny binárně. Přesněji to tedy znamená, že zakoupená služba je označena číslicí 1 a nezakoupená služba je označena číslicí 0. Na tuto datovou instanci je aplikován algoritmus asociačních pravidel a jsou hledány vzájemné zákonitosti mezi jednotlivými službami.



Obr. 5 Znáznornění modelu na určení asociačních pravidel v programu RapidMiner (zdroj: autor)

Na obrázku číslo 5 je uveden model, jenž zajišťuje rozbor druhé tabulky, a to na základě použití asociačních pravidel. Ikona *Read Excel* slouží pro import tabulky číslo 2 do programu

RapidMiner. Na načtená data je navázán operátor *Replace missing values*, který v datové instanci vhodně nahradí chybějící proměnné. V našem případě je datová instance kompletní, avšak v případě velkých databází nemusí být zřejmé, zda nějaký atribut neobsahuje chybějící hodnotu proměnné. Z tohoto důvodu, jsem se rozhodl tento operátor zařadit a připravit si tak celkový model na následující rozbor databáze firmy UVXA. *Replace missing values* operátor je následně propojen na *Select attributes*, díky kterému je možno vybrat atributy potřebné pro výpočet. Další propojení se váže k operátoru Discretize (Discretize by Frequency). Tento operátor zajišťuje diskretizaci reálných atributů do nastaveného počtu kategorií. Proces diskretizace je nutnou součástí metody asociačních pravidel, a to kvůli tomu, že RapidMiner je schopný pracovat pouze s binárními hodnotami proměnných. V případě zkušební datové instance je zadání proměnných správné, použití tohoto operátoru má ale stejný význam, jako nasazení operátoru *Replace missing values*. Po diskretizování atributů pokračují data do operátoru *Nominal to Binominal*, jenž případně zajišťuje převod proměnných a celkového atributu na atribut a proměnné binární.

Dále je navázán operátor *FP-Growth (Frequent Pattern Growth)*, jenž vyhledá množinu atributů jednotlivých frekventovaných vzorů, které splňují určitý kritérium. Kritériem v případě tohoto atributu je minimální podpora (*support*), vyjadřovaná v zásadě procentuálně v závislosti na celkovém počtu objektů v databázi. Hodnota minimální podpory musí vždy náležet intervalu $<0,1>$. Konečným atributem ve výše zobrazeném modelu je *Create Association Rules*. Ten nejdříve nalezne množinu frekventovaných vzorů s nejmenší hodnotou podpory z předchozího operátoru a následně začne z této množiny generovat asociační pravidla. Vytváření asociačních pravidel je podmíněno kritériem, které vyhledávání podmiňuje. Možností výběru kritérií je hned několik. Ve všech užitých případech užívání asociačních pravidel v této práci je však užívána minimální spolehlivost, tudíž v nastavení operátoru nastavíme parametr *criterion* na hodnotu *confidence*, což nám následně umožní zadávat hodnoty do parametru *min confidence* na intervalu $<0,1>$.

Pokud je celý model pro zjištění asociačních pravidel správně nastaven, po spuštění se nám zobrazí veškerá nalezená pravidla, která splňují předem zadané požadavky z jednotlivých operátorů. Vzhledem k počtu spojů, které jsou v modelu uvedeny, výstupem jsou hodnoty z tří operátorů. V tomto konkrétním případě nás nejvíce zajímají výsledky zkoumání operátorů *FP-Growth* a *Create Association Rules*. Operátor *Discretize (Discretize by Frequency)* nám zde přináší pouze informace o počtu kategorií. Výsledky dvou klíčových operátorů jsou přehledně zaneseny do následujících tabulek.

Size ↑	Support	Item 1	Item 2	Item 3
1	1.000	Digitální televize		
1	0.600	Internetové připojení		
1	0.500	Analogová televize		
2	0.600	Digitální televize	Internetové připojení	
2	0.500	Digitální televize	Analogová televize	
2	0.500	Internetové připojení	Analogová televize	
3	0.500	Digitální televize	Internetové připojení	Analogová televize

Obr. 6 výstup operátoru *FP-Growth (Frequent Pattern Growth)* (zdroj: autor)

Výstupem operátoru *FP-Growth (Frequent Pattern Growth)* je tabulka, ve které jsou uvedeny jednotlivé četnosti nabízených produktů v závislosti na jejich prodeji. Počet zkoumaných produktů nám popisuje první sloupec (*Size*). Ve druhém sloupci se nalézají hodnoty podpory (*Support*), jež lze prezentovat jako procentuální zastoupení pořízení jednotlivé služby či kombinaci služeb. V následujících třech sloupcích (*Item 1*, *Item 2*, *Item 3*) jsou uvedeny samostatné služby a kombinace služeb, pro které byly stanoveny konkrétní hodnoty podpory.

První tři řádky ukazují hodnoty pro jednotlivé služby (*digitální televize*, *internetové připojení* a *analogová televize*). U služby *digitální televize* je na první pohled patrné, že jí má zakoupeno 100% zákazníků z analyzované databáze. U služby *internetové připojení* je to pak 60% a službu *analogová televize* má pořízeno 50% zákazníků. Důležitějším prvkem jsou kombinace služeb. Pokud se zaměříme na kombinace dvou služeb zjistíme, že nejoblíbenější kombinací je pro zákazníky duo *digitální televize* a *internetové připojení*. Kombinaci všech služeb využívá 50% zákazníků.

No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain
1	Digitální televize	Analogová televize	0.500	0.500	0.750	-0.500
2	Digitální televize	Internetové připojení, Analogová televize	0.500	0.500	0.750	-0.500
3	Digitální televize	Internetové připojení	0.600	0.600	0.800	-0.400
4	Internetové připojení	Analogová televize	0.500	0.833	0.938	-0.100
5	Internetové připojení	Digitální televize, Analogová televize	0.500	0.833	0.938	-0.100
6	Digitální televize, Internetové připojení	Analogová televize	0.500	0.833	0.938	-0.100
7	Internetové připojení	Digitální televize	0.600	1	1	0
8	Analogová televize	Digitální televize	0.500	1	1	0
9	Analogová televize	Internetové připojení	0.500	1	1	0
10	Analogová televize	Digitální televize, Internetové připojení	0.500	1	1	0
11	Digitální televize, Analogová televize	Internetové připojení	0.500	1	1	0
12	Internetové připojení, Analogová televize	Digitální televize	0.500	1	1	0

Obr. 7 výstup operátoru Create Association Rules (zdroj: autor)

Obrázek číslo 7 zobrazuje výstup konečného operátoru Create Association Rules, sloužícího k finálnímu vytvoření asociačních pravidel. Výstup z posledního operátoru je uveden ve formě přehledné tabulky a přináší hned několik různých asociačních pravidel. První sloupec slouží pouze jako ID u jednotlivých asociačních pravidel. Druhý a třetí sloupec jsou spolu velmi úzce spjaty a ve své podstatě díky nim lze interpretovat, jak se jednotlivá asociační pravidla chovají. Prvním z dvojice je sloupec s označením atributu *Premises*, zastávající v konstrukci asociačních pravidel část IF. Druhým z páru je sloupec označený atributem *Conclusion* obsahující proměnné, které reprezentují druhou část konstrukce asociačních pravidel a to sice část THEN.

Konstrukce IF-THEN (Premises-Conclusion) je, jak již bylo zmíněno, nejdůležitější konstrukcí asociačních pravidel. Konkrétní konstrukci můžeme demonstrovat například u asociačního pravidla číslo 3, které obsahuje hodnoty (digitální televize; internetové připojení) a hodnoty (support 0.500; confidence 0.833). Pokud bychom tento vztah chtěli zapsat v souvislé větě, zněl by následovně. Pokud bude uživatel mít zájem o službu digitální televize, můžeme s pravděpodobností 0.833 předpokládat, že bude mít zároveň zájem o službu internetové připojení. Hodnota podpory (support) zde určuje četnost výskytu takového případu v závislosti na veškerých analyzovaných případech.

Počet asociačních pravidel, jež můžeme získat při aplikování procesu na databázi, může být naprosto rozdílný a jejich množství záleží například na minimálním určení podpory (support) a jistoty (confidence). Při nastavení nižší hodnoty u atributu support dojde k vygenerování mnohem většího množství asociačních pravidel. Stejně tak je to i

s parametrem (minimal confidence), ve kterém si můžeme určit jaká minimální pravděpodobnost nám bude limitovat vypsání asociačního pravidla. Jako poslední fakt při popisu této tabulky bych rád zmínil poslední řádek neboli dvanácté asociační pravidlo. V části *premises* se nám vyskytují dva produkty, a to sice *internetové připojení* a *analogová televize*. Podmínkou jsou tedy v tomto případě dvě služby a poslední asociační pravidlo lze interpretovat následovně. Pokud si zákazník pořídí internetové připojení a zároveň k tomu službu analogové televize, můžeme s absolutní jistotou (100%) předpokládat, že si pořídí i digitální televizi.

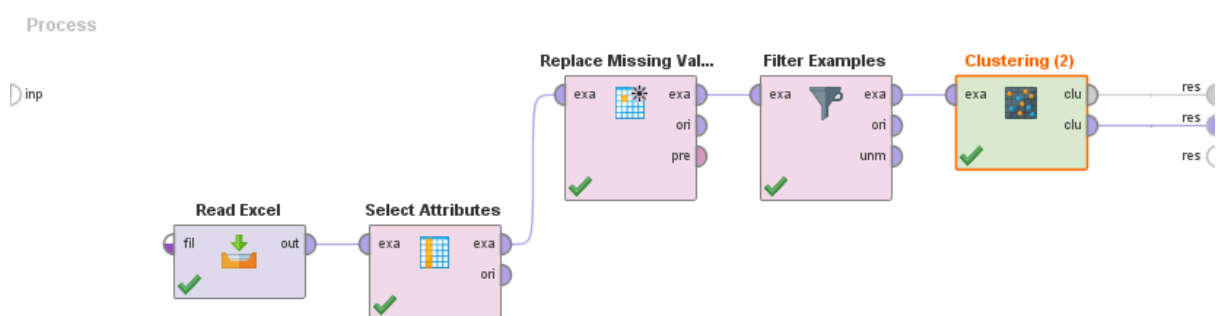
4.4.3 Aglomerativní shlukování

Další metodou, pomocí které lze analyzovat různé datové sady, je shluková analýza. Shluková analýza je často také nazývána clusterová analýza a ve své podstatě funguje na principu klasifikace každého objektu databáze do určitých shluků. Každá ze skupin shluků má své vlastnosti, jež jsou vždy výrazně odlišné od vlastností jiné skupiny. Na základě jednotlivých vlastností shluků jsou do skupin přiřazovány objekty z databáze, které mají vždy největší vzájemnou podobnost a tím pádem vznikají shluky podobných objektů. Existují dva různé typy členění shlukování, zaměřené povětšinou na různé cíle. Prvním typem je označován jako hierarchické shlukování, vytvářející rozsáhlý systém podmnožin, kde jejich vzájemným průnikem může být buďto prázdná množina či jeden ze shluků (Berka, 2005). Do sekce hierarchického shlukování patří i zmíněné aglomerativní shlukování, jež funguje na principu spojování shluků obsahujících jeden člen. Hlavním problémem aglomerativního shlukování je fakt, že jejich použití je podmíněno menšími datovými sadami, což je v případě této bakalářské práce a následně analyzované databáze firmy UVXA vyloučeno. Pro nastínění fungování jsem vytvořil datovou sadu, na kterou je možno aglomerativní shlukování aplikovat a následně popsat jeho výstupy.

Zákazník	Kombinace produktů	Cena před slevou	Sleva %	Konečná cena	Věk
Spotřebitel A	DA	550	2	539	21
Spotřebitel B	DIV	850	6	799	35
Spotřebitel C	I	350	0	350	25
Spotřebitel D	ADIV	1000	15	850	28
Spotřebitel E	IV	450	1	445,5	42
Spotřebitel F	AV	250	0	250	59
Spotřebitel G	DV	500	6	470	47
Spotřebitel H	ADIV	1000	15	850	61
Spotřebitel I	AIV	600	5	570	18
Spotřebitel J	ADV	650	8	598	21

Tab. 3 Ukázková data pro aglomerativní shlukování (zdroj: autor)

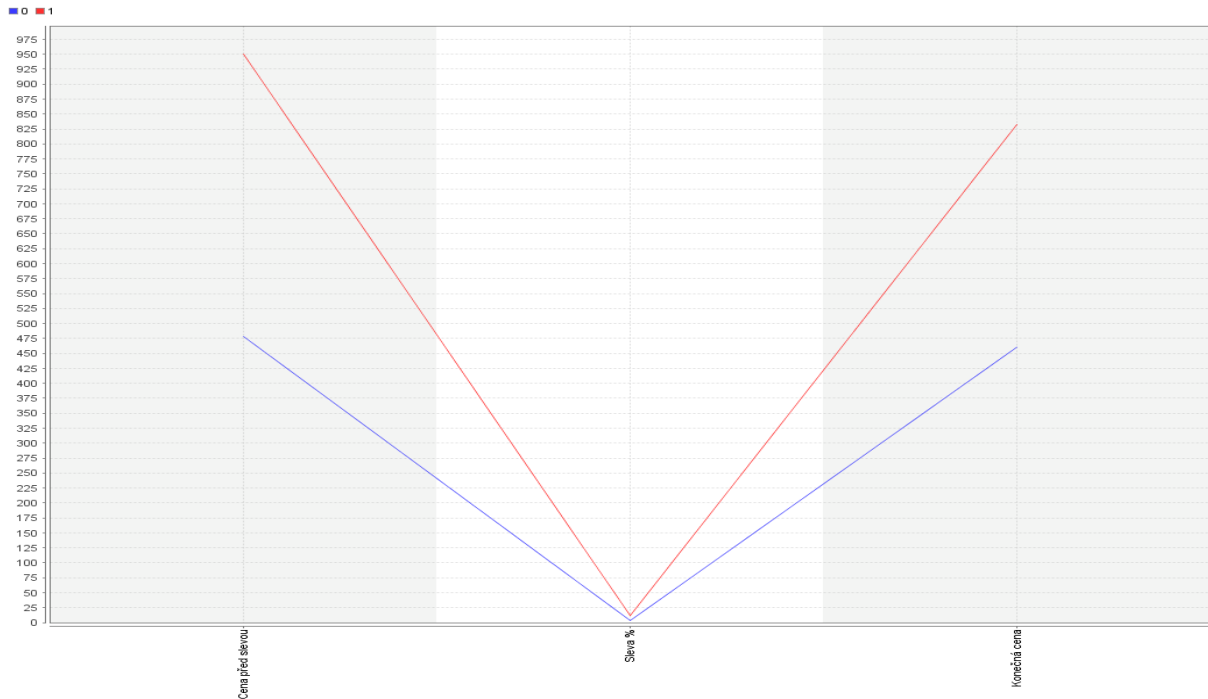
Tabulka číslo 3 obsahuje potenciaální údaje o kombinaci zakoupených služeb, ceně a věku. Ke každému zákazníkovi jsou uvedeny pořízené produkty a následně jejich cena před slevou. Ve čtvrtém sloupci nalezneme informace o procentuální výši slevy, jež byla zákazníkovi poskytnuta. Výše této slevy je závislá na kombinaci pořízených produktů, přičemž sleva není poskytována na samostatné pořízení služby či při pořízení dvou nejlevnějších služeb. Data, obsažená v této tabulce, mají pouze informativní charakter a nijak nekorrespondují s údaji z databáze firmy UVXA. Celá datová instance uvedená v tabulce číslo 3 slouží pouze k demonstraci fungování aglomerativního shlukování.



Obr. 8 Aplikace aglomerativního shlukování v programu RapidMiner (zdroj: autor)

První tři operátory již byly popsány například u obrázku číslo 5 a v modelu z obrázku číslo 8 mají naprosto stejné využití. Novým operátorem je *Filter Examples*, zajišťující vyřazení záznamů, které jsou pro správné fungování aglomerativního shlukování závadné. Odstranění závadných záznamů je provedeno pomocí parametru

condition class, jenž je nastaven na hodnotu *no_missing_attributes*. Poslední vazba vede k operátoru Clustering, konkrétně (Agglomerative clustering), díky kterému je možné provést analýzu aglomerativního shlukování.



Obr. 9 Grafický výstup z modelu pro určení aglomerativního shlukování (zdroj: autor)

Na grafu modelu aglomerativního shlukování je zřetelně vidět, že zkoumaná data byla rozdělena do dvou hlavních klastrů neboli podobných tříd. Klastry jsou na grafu prezentovány modrou a červenou křivkou, které představují shluk dat s podobnými hodnotami. Hlavními atributy, které ovlivňují rozdělení dat do klastrů jsou atributy *cena před slevou* a *konečná cena*. Tento fakt je na výše uvedeném grafickém znázorněn velkou rozdílnou hodnotou u klíčových atributů, a to mezi křivkou klastru 0 a křivkou klastru 1.

5. Praktická část

5.1 Databáze zákazníků firmy UVXA

Jak již bylo několikrát zmíněno, pro správné využití data miningových technik je nezbytné velké množství dat. Tento fakt v plném rozsahu splňuje databáze firmy UVXA, která mi byla poskytnuta pro studijní účely a plně koresponduje s reálným postavením firmy na trhu telekomunikačních služeb v České republice. Datová sada obsahuje veškeré získané údaje o zákaznících, které má firma k dispozici. Celkový počet zákazníků firmy je 562 842 a u každého zákazníka je uvedeno 42 informací, které mohou sloužit k potencionálnímu zvýšení prodeje služeb či k získání nových zákazníků. Firma UVXA uchovává informace o svých zákaznících již po dlouhou dobu a podle informací, které jsem získal z marketingového oddělení firmy, jsou jim techniky data miningu dobře známé. Poskytnutá databáze odráží reálné informace o zákaznících za rok 2016 a to konkrétně do začátku prosince roku 2016, kdy mi byla databáze poskytnuta. Jak již bylo zmíněno na začátku práce, při převzetí dat jsem podepsal písemnou dohodu s poskytovatelem, kde jsem se zavázal k použití pouze pro studijní účely a také k neuvedení názvu společnosti. Databáze totiž obsahuje velmi citlivá data, zejména pro společnost UVXA, které by mohly ohrozit její postavení na trhu.

Označení atributu	Popis atributu	Označení atributu	Popis atributu
CUST_ID	ID zákazníka	REGION_NAME	Název Kraje
CUST_GENDER	Pohlaví zákazníka 0/1 MUŽ/ŽENA	POP_TOTAL	Celková populace města
CUST_AGE	Věk zákazníka	POP_MALE	Populace mužů ve městě
CITY_ID	ID města zákazníka	POP_FEMALE	Populace žen ve městě
CITY_NAME	Název města	AGE_AVG	Průměrný věk ve městě
DISTRICT_ID	ID okresu	AGE_MALE	Průměrný věk mužů ve městě
DISTRICT_NAME	Název okresu	AGE_FEMALE	Průměrný věk žen ve městě
REGION_ID	ID kraje	FLAG_ATV	Zákazník má pořízenou analogovou televizi
FLAG_DTV	Zákazník má pořízenou digitální televizi	A_PRICE_BASE	Základní cena analogové TV
FLAG_INET	Zákazník má pořízený internet	D_PRICE_BASE	Základní cena digitální TV
FLAG_VOICE	Zákazník má pořízenou hlasovou službu	I_PRICE_BASE	Základní cena internetu
NO_OF_BL	Počet pořízených služeb	V_PRICE_BASE	Základní cena hlasové služby
COMBINATION	Kombinace služeb	PRICE_LEVEL	Základní cenová úroveň
PRODUCT_ATV_ID	ID analogové služby	PRICE_LEVEL_DISCOUNT	Sleva na základní cenu
PRODUCT_DTV_ID	ID digitální služby	PRODUCT_EXTRA_ID	ID extra služby
PRODUCT_INET_ID	ID internetové služby	PRODUCT_EXTRA_NAME	Název extra služby
PRODUCT_VOICE_ID	ID hlasové služby	PRODUCT_EXTRA_PRICE	Cena extra služby
PRODUCT_ATV_NAME	Název analogové služby	PRICE_TOTAL_w.o_EXTRA	Celková cena bez extra služby

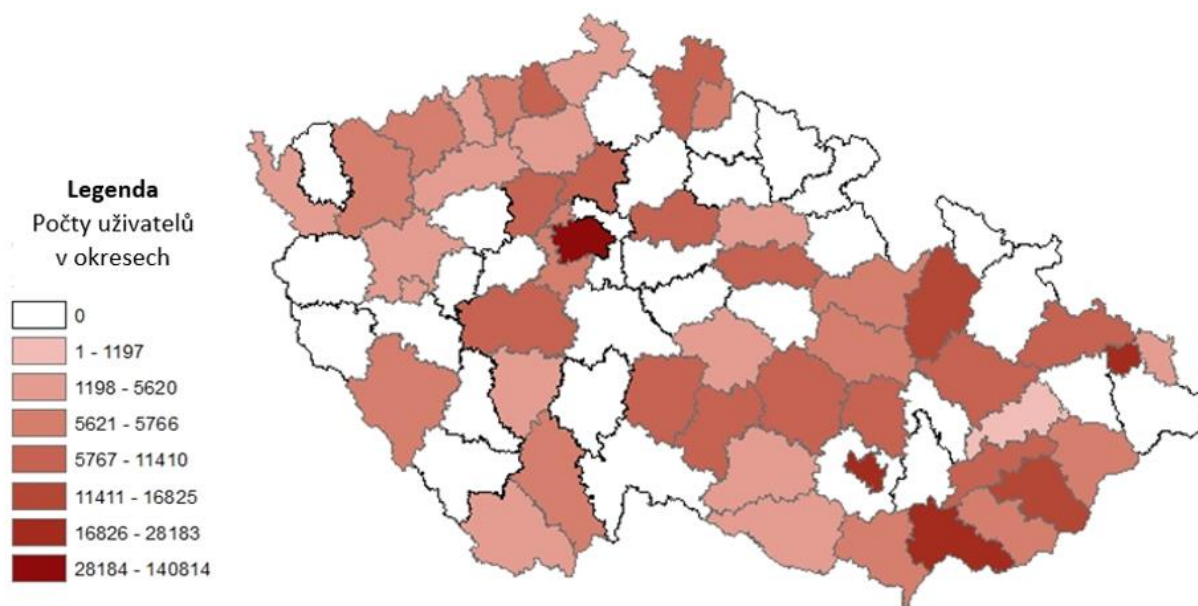
PRODUCT_DTV__NAME	Název digitální služby	PRICE_TOTAL	Celková cena před konečnou slevou
PRODUCT_INET_NAME	Název internetové služby	DISCOUNT_ (%)	Výše slevy
PRODUCT_VOICE_NAME	Název hlasové služby	PRICE_LEVEL	Konečná cena po slevě

Tab. 4 Popis atributů databáze firmy UVXA (zdroj: autor)

Tabulka číslo 4 zobrazuje kompletní výčet služeb a údajů, které má firma UVXA k dispozici. Některé z atributů obsahují chybějící hodnoty, a to například v případech, kdy si zákazník nepořídí službu analogové televize, která je pak v databázi následně popisována. V takovýchto případech je buňka v databázi ponechána s prázdnou hodnotou. Vzhledem k faktu, že databáze obsahuje velké množství informací, rozhodl jsem se pro následné analyzování databáze firmy použít jen některé, jen jsou vždy u konkrétního druhu rozboru uvedeny.

Z důvodu lepšího pochopení působení firmy UVXA na českém trhu jsem se nejprve rozhodl zaměřit na celkové množstevní rozložení zákazníků v České republice. Pro zvýšení efektivity a lepší grafické znázornění jsem použil software ArcGIS, který je součástí Geografického Informačního Systému (GIS). Ten slouží primárně k analýze dat, na základě které, je následně schopen provést přehlednou vizualizaci. V případě této bakalářské práce jsou veškeré výstupy analýz dat graficky znázorněné na mapě České republiky. V následující kapitole je uvedeno několik výstupů z programu GIS, které znázorňují například rozložení zákazníků firmy UVXA či celkové zisky plynoucí z prodeje služeb v jednotlivých krajích. Veškeré výstupy jsou nadále použity pro následnou predikci skupin zákazníků firmy UVXA.

5.2 Grafická analýza okresů databáze firmy UVXA v prostředí ArcGIS



Obr. 10 Grafické znázornění počtu uživatelů na mapě České republiky (zdroj: autor)

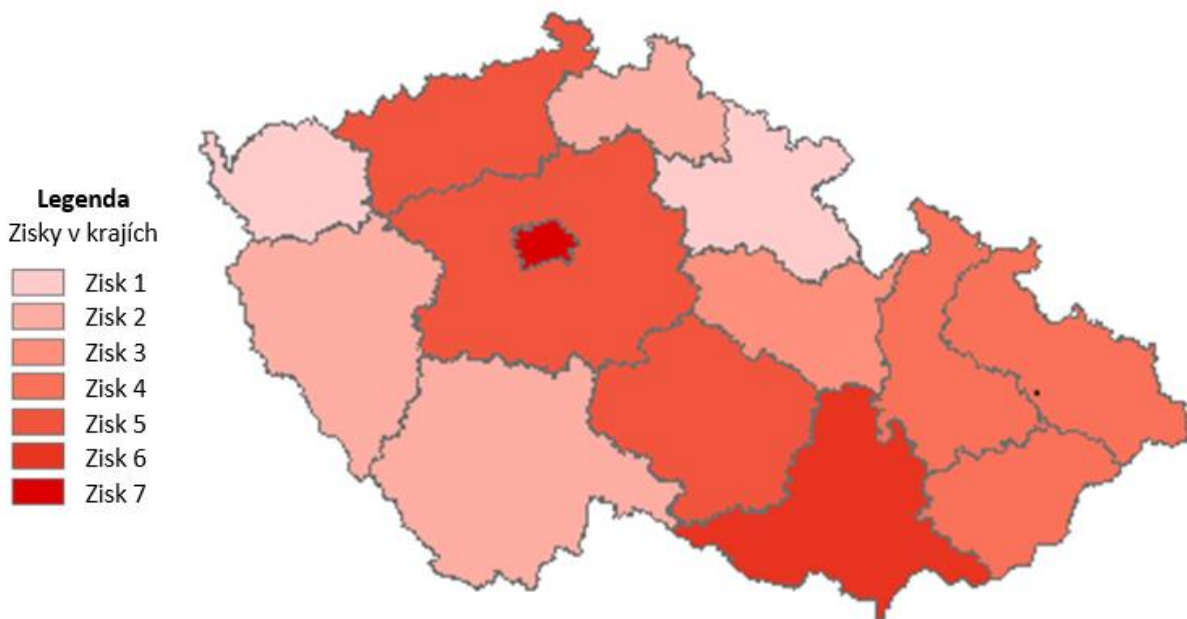
Jelikož databáze firmy UVXA neudává přesný počet zákazníků u jednotlivých okresů v samostatném sloupci, bylo nutné nejdříve provést náročnou datovou úpravu, při které jsem shromáždil konkrétní počty uživatelů v okresech. Po shromáždění počtů zákazníků v jednotlivých okresech jsem provedl ruční import dat do Geografického Informačního Systému (GIS) a to sice jako parametry okresů. Následně jsem pomocí těchto načtených hodnot nechal zobrazit grafické rozložení zákazníků firmy UVXA na mapě České republiky (obrázek číslo 10).

Na základě mapy je na první pohled patrné, že firma UVXA nemá zastoupení ve všech okresech. Bílé oblasti prezentují okresy České republiky, ve kterých nemá firma ani jednoho zákazníka. Hlavním důvodem absence uživatelů v některých okresech je rozložení páteční sítě firmy UVXA, která se ve své podstatě ubírá třemi směry. Hlavním uzlem je hlavní město Praha, ze kterého se páteční síť rozpíná do třech směrů. Prvním směrem je jižní napojení a prochází přes okresy Příbram, Písek, České Budějovice a Český Krumlov. Druhý směr je orientovaný na východ, a to sice přes Nymburk, Hradec Králové, Pardubice a Svitavy, kde se následně rozděluje na dva uzly. První uzel je orientovaný do Moravskoslezského kraje, konkrétně do Ostravy a dále do Polska. Druhý uzel pokračuje směrem do Zlínského kraje a Jihomoravského, kde dále pokračuje do zahraničních zemí. Poslední napojení vychází z Prahy

a je orientováno na severozápad a západ. Dalším důvodem absence uživatelů je tvrdá konkurence v odvětví poskytování telekomunikačních služeb, kdy se každý provider snaží své stávající zákazníky hájit a z tohoto důvodu často dochází k zabraňování vstupu do okresů ostatním poskytovatelům.

Závěrem popisu obrázku číslo 10 bych chtěl zmínit fakt, že zobrazená množina uživatelů vychází pouze ze zákazníků, kteří mají fyzicky podepsanou smlouvu přímo s firmou UVXA. Firma UVXA má totiž ještě několik svých partnerů, prostřednictvím kterých své služby poskytuje. Informace o těchto uživatelích však nelze získat, jelikož jsou vždy ve vlastnictví dealerů služeb a ti nemají do databáze firmy UVXA přístup. Vzhledem k ochraně společnosti UVXA nemohu uvést, o jaké dealery se jedná. Pokud bychom tedy měli veškeré databáze od dealerů je možné, že by bylo zastoupení zákazníků v okresech odlišné. Obrázek číslo 10 tedy odráží reálnou situaci zastoupení uživatelů za rok 2016 pouze u zákazníků majících zakoupený produkt přímo u firmy UVXA.

5.3 Grafická analýza krajů databáze firmy UVXA v prostředí ArcGIS



Obr. 11 Grafické znázornění zisků firmy UVXA v krajích na mapě České republiky (zdroj: autor)

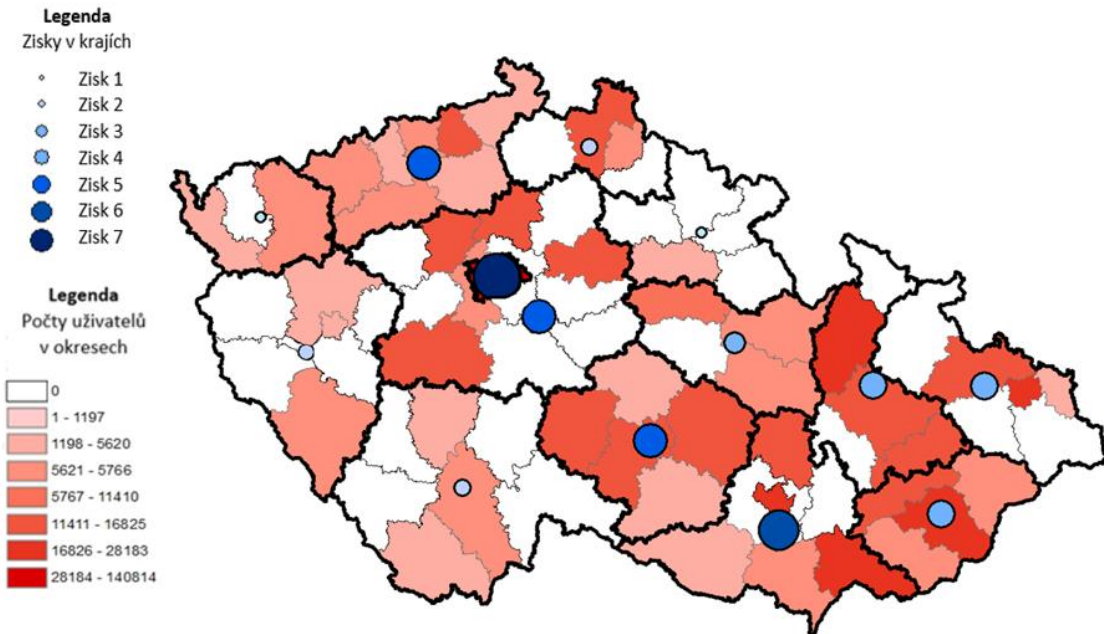
Úvodem popisu obrázek číslo 11 je nutné osvětlit informace obsažené v legendě. Popisky legendy (Zisk 1, Zisk 2, Zisk 3 atd.) jsou uvedeny v této formě z praktického důvodu. Grafické znázornění je založeno na reálných hodnotách, které jsou uvedeny v řádu milionů

českých korun. Vzhledem k uchování citlivých informací o firmě UVXA jsem se rozhodl použít tuto formu legendy, což však nijak neovlivňuje pravdivostní hodnotu zobrazené mapy.

Nejvyšší zisky přináší firmě UVXA hlavní město Praha. Důvodem je bezesporu nejpočetnější zastoupení zákazníků v hlavním městě a také velká četnost kombinací služeb a tím i vyšší zisky z jednotlivého zákazníka. Zisky v hlavním městě Praha například více jak patnáctinásobně přesahují příjmy v Královehradeckém kraji. Na základě informací z obrázku číslo 11 můžeme jednoznačně určit, v jakých krajích je firma UVXA úspěšná či nikoliv. Druhým kraje přinášejícím firmě UVXA vysoké zisky je kraj Jihomoravský. I přes fakt, že tento kraj je v ziscích druhý, museli bychom vynásobit hodnotu zisku Jihomoravského kraje více než dvakrát, abychom se dostali na hodnotu hlavního města Prahy. Ostatní kraje lze rozčlenit do čtyř ekonomických skupin.

Do první skupiny krajů lze zahrnout kraje Ústecký, Středočeský a Jihlavský. V těchto krajích jsou hodnoty zisků velmi podobné a odlišnost v podobě násobku je zanedbatelná. Druhá skupina obsahuje Moravskoslezský, Olomoucký, Zlínský a potažmo i Pardubický kraj. Z druhé skupiny krajů přináší nejnižší zisky kraj Pardubický a nejvyšší zisky kraj Moravskoslezský, jehož výše násobku hodnoty oproti Pardubickému kraji dosahuje hodnoty 1,7. Zbývající dva kraje (Olomoucký a Zlínský) mají hodnoty zisku velmi podobné. Třetí skupina je vytvořena z krajů Jihočeského, Plzeňského a Libereckého. U těchto krajů jsou hodnoty také velmi podobné a násobky hodnot jsou zanedbatelné. Poslední skupinu lze označit za skupinu kritickou pro firmu UVXA. Čtvrtou skupinu tvoří kraje Královehradecký a Karlovarský. U těchto krajů firma dosahuje nejnižších zisků a tím pádem nepřináší firmě UVXA takový prospěch jako ostatní kraje.

Vzhledem k tématu této bakalářské práce (Získávání zákazníků E-commerce: využití metod data mining), jsem se rozhodl pro analyzování dvou nejslabších krajů z hlediska zisků, a to sice krajů Královehradeckého a Karlovarského. Na následujících stránkách je proveden rozbor každého z dvou uvedených krajů pomocí data miningových technik, díky kterým jsou získány velmi hodnotné informace. Na základě těchto informací se snažím analyzovat chování zákazníků v ekonomicky slabých krajích a hledat adekvátní způsob, kterým by bylo možné získat v problematických krajích nové zákazníky.



Obr. 12 Grafické propojení počtu uživatelů v okresech a zisků v krajích (zdroj: autor)

Mapa, zobrazená na obrázku číslo 11, je posledním výstupem z prostředí ArcGIS a tedy celého Geografického Informačního Systému (GIS). Ve své podstatě je to kombinace dvou předešlých map a slouží k definitivnímu pochopení situace firmy UVXA na českém trhu telekomunikačních služeb. Jak již bylo zmíněno, nejmenší zisky přináší pro firmu UVXA pětice krajů Královéhradecký, Karlovarský, Jihočeský, Plzeňský a Liberecký (seřazeno vzestupně dle výše zisků). Nejhorší situace je v Královéhradeckém kraji, kde mají uživatelé zakoupeny služby pouze v jednom okrese. V Karlovarském kraji, který zastupuje druhou pozici nejméně prosperujícího regionu, je zastoupení uživatelů v okresech z 66,66 %, což ale stále nestačí k dosahování dostatečného zisku, kterým by byl Karlovarský kraj schopen konkurovat ostatním krajům. U zbylé trojici krajů (Jihočeský, Plzeňský, Liberecký) nebude prováděna podrobná analýza krajů pomocí metod data miningu, jelikož přináší lepší zisky než Královéhradecký a Karlovarský.

5.4 Analýza Královehradeckého kraje

Královehradecký kraj vyšel v porovnání krajů České republiky jako jednoznačně nejslabší a z tohoto důvodu jej uvádím jako první kraj pro analyzování. Pokud se zaměříme na zastoupení uživatelů v jednotlivých okresech Královehradeckého kraje zjistíme že jediné zastoupení uživatelů je v okrese Hradec Králové. Ostatní okresy (Jičín, Trutnov, Náchod a Rychnov nad Kněžnou) zůstávají bez zastoupení. V Hradci králové má firma UVXA 3604 zákazníků, z nichž každý má jiné preference nákupu služeb. Pro analýzu Královehradeckého kraje a následné určení nejlepší kombinace prodeje služeb použijí asociační pravidla a rozhodovací stromy. Na základě získaných informací doporučím firmě UVXA kombinaci služeb, které by mohly potencionálně vést ke zvýšení počtu zákazníků v nejméně ziskovém kraji České republiky.

Size	Support	Item 1	Item 2	Item 3
1	0.569	FLAG_INET		
1	0.278	FLAG_ATV		
1	0.252	FLAG_VOICE		
2	0.210	FLAG_INET	FLAG_ATV	
2	0.145	FLAG_INET	FLAG_VOICE	
2	0.071	FLAG_ATV	FLAG_VOICE	
3	0.055	FLAG_INET	FLAG_ATV	FLAG_VOICE

Obr. 13 Výstup z operátoru FP-Growth pro databázi Královehradeckého kraje (zdroj: autor)

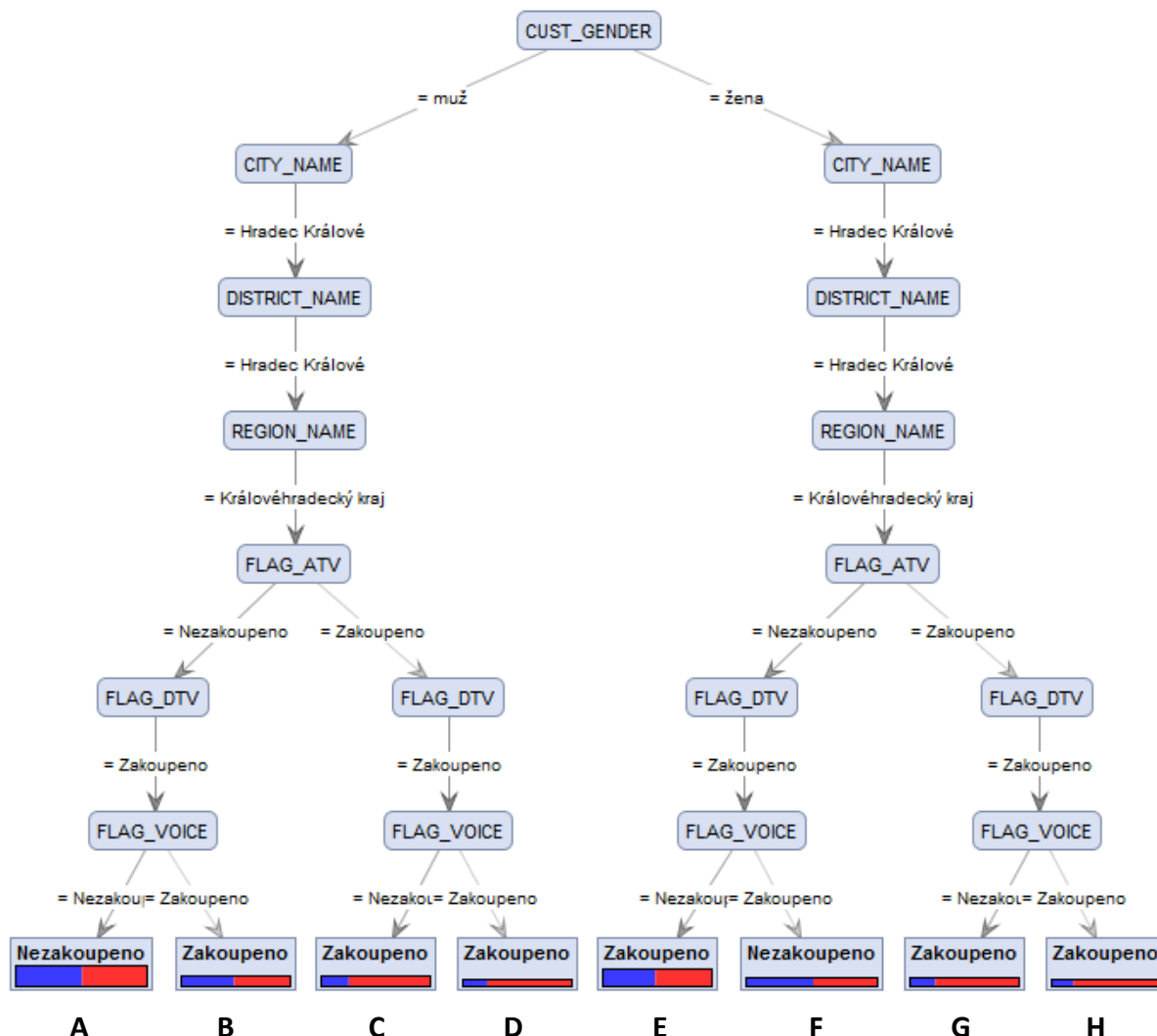
Obrázek číslo 12 zobrazuje tabulku výstupu z operátoru FP-Growth. První sloupec označuje počet kombinací služeb, které se v databázi vyskytují jako zakoupené. Z možností případů nám vypadl atribut FLAG_DTV, a to kvůli tomu, že tuto službu mají zakoupenou všichni zákazníci. Četnost případů zakoupení jednotlivých služeb je prezentována sloupcem *Support* a odráží reálné zastoupení nákupů služeb u zákazníků. Pokud nebereme v potaz službu FLAG_DTV, kterou má zakoupeno 100 % zákazníků, pak je nejoblíbenější službou pro zákazníky Královehradeckého kraje služba FLAG_INET, jež je zastoupena v 56,9 % případů. V kombinaci dvou produktů dominuje kombinace FLAG_INET a FLAG_ATV, které jsou u zákazníků oblíbené z 25,2 % případů. V případě trojice služeb je procentuální zastoupení velmi malé a to v 5,5%

Size	Support	Item 1	Item 2	Item 3	Total average price	Zisky pro firmu UVXA
1	0.569	FLAG_INET			369	Zisk 1
1	0.278	FLAG_ATV			143	Zisk 5
1	0.252	FLAG_VOICE			216	Zisk 4
2	0.210	FLAG_INET	FLAG_ATV		512	Zisk 2
2	0.145	FLAG_INET	FLAG_VOICE		585	Zisk 3
2	0.071	FLAG_ATV	FLAG_VOICE		359	Zisk 6
3	0.055	FLAG_INET	FLAG_ATV	FLAG_VOICE	728	Zisk 5

Tab. 5 Četnosti služeb v Královehradeckém kraji a jejich nacenění (zdroj: autor)

Tabulka číslo 5 vychází z výstupu operátoru FP_Growth, avšak je doplněná o informace o cenách za jednotlivé služby. Veškeré ceny u produktů vychází z průměrovaných hodnot z databáze zákazníků Královehradeckého kraje. Každý zákazník si totiž může objednat různé varianty u jednotlivých služeb, u kterých je cena odlišná. Průměrné ceny zákazníků Královehradeckého kraje jsou uvedeny ve sloupci *total average price* a reprezentují jak ceny jednotlivých služeb, tak i kombinace dvou a více služeb. Sloupec *zisky pro firmu UVXA* představují výši zisků, které služba či kombinace služeb firmě UVXA přináší. Údaje (Zisk 1 – Zisk 6) jsou založené na reálných hodnotách, ale z důvodu ochrany společnosti UVXA byly uvedeny v této formě. Nejvyšší zisky přináší firmě UVXA v Královehradeckém kraji služba FLAG_INET a z tohoto důvodu se v následujícím rozhodovacím stromu zaměřím na tuto službu.

5.4.1 Rozhodovací strom Královehradecký kraj



Obr. 14 Rozhodovací strom pro službu FLAG_INET v Královehradeckém kraji (zdroj: autor)

Výše uvedený obrázek rozhodovacího stromu zobrazuje zákonitosti, které vedou zákazníka ke koupi služby FLAG_INET. První rozdělení probíhá mezi muže a ženy a následný postup hledání zákonitostí mezi daty je pro obě pohlaví stejný. Nejdříve jsou zkoumány atributy *city_name*, *district_name* a *region_name*, které jsou zde uvedeny pro případ, že by zákazníci pocházeli z různých měst, okresů či krajů. První členění mužů a žen přichází u služby FLAG_ATV, kde se řeší, jestli má zákazník službu pořízenou či nikoliv. Atribut FLAG_ATV rozděluje muže i ženy do skupin, na základě pořízení či nepořízení služby. O službě FLAG_DTV víme, že jí mají zakoupenou všichni zákazníci, tudíž není limitujícím atributem.

Poslední selekce zákazníků u mužů a žen probíhá u služby FLAG_VOICE, jejíž hodnoty rozdělí zákazníky u služby FLAG_INET do konečných skupin. Tento rozhodovací strom uznává jako rozhodující kritérium pro určení hodnoty labelu (tím pádem i skupin zákazníků) počet nezakoupených/zakoupených služeb. Z tohoto důvodu například označuje pořízení služby FLAG_INET u *množiny A* jako celkově nezakoupenou, i když obsahuje největší počet zakoupení od uživatelů. Pro detailnější popis rozhodovacího stromu z obrázku číslo 13 je použita následující tabulka, která popisuje hodnoty jednotlivých množin zákazníků a potencionální preference firmy UVXA.

Označení množiny	Počet nezakoupení	Počet zakoupení	Pozice dle zisku	Procentuální zakoupení	Pozice dle procent
Muži					
Množina A	561	536	1	48,86	4
Množina B	117	183	3	61,00	3
Množina C	109	307	2	73,80	2
Množina D	33	109	4	76,76	1
Ženy					
Množina E	419	436	1	50,99	3
Množina F	151	140	3	48,11	4
Množina G	79	249	2	75,91	2
Množina H	24	91	4	79,13	1

Tab. 6 Reprezentace číselných hodnot z rozhodovacího stromu pro Královehradecký kraj

Množiny obsahují různé počty zákazníků, přičemž u mužů je nejpočetněji zastoupena cesta (FLAG_ATV/NEZAKOUPENO -> FLAG_DTV/ZAKOUPENO -> FLAG_VOICE/NEZAKOUPENO -> FLAG_INET/ZAKOUPENO). V tomto případě cesty je výstupní množina označena písmenem A (*Množina A*), přičemž stěžejní informací je pro nás počet zakoupení služby FLAG_INET čítající 536 zákazníků. V případě žen je primárně zastoupena stejná cesta a kombinace produktů jako u mužů (FLAG_ATV/NEZAKOUPENO -> FLAG_DTV/ZAKOUPENO -> FLAG_VOICE/NEZAKOUPENO -> FLAG_INET/ZAKOUPENO), kterou zobrazuje písmeno E (*Množina E*). I v tomto případě dosahuje tato cesta nejvyššího počtu zakoupené služby FLAG_INET u zákazníků, tentokrát u žen. Tyto množiny lze tedy označit jako nejprínosnější pro firmu UVXA, v případě, že budeme brát jako limitující faktor výši zisků, které daná množina uživatelů firmě přináší. Sloupec *pozice dle zisku* tedy udává pořadí množin, řazených čistě podle výše zisků, v závislosti na počtu zakoupení služby FLAG_INET. Hodnocení prosperity množin je však možné zkoumat i z jiného hlediska. Jako další kritérium jsem zvolil

procentuální vyjádření vztahu mezi zakoupenou a nezakoupenou službou. Tento vztah reprezentuje sloupec *procentuální zakoupení*, který uvádí, kolik procent zákazníků dané množiny si zakoupilo službu FLAG_INET. Pokud by pro firmu UVXA byla primární výše procentuálního úspěchu zakoupení cílové služby, pak jako nejlepší řešení vychází cesta, vedoucí k množině D u mužů a k množině H u žen.

5.4.2 Asociační pravidla Královehradecký kraj

No.	Premises	Conclusion	Support	Confidence
5	FLAG_INET, FLAG_ATV	FLAG_VOICE	0.055	0.265
6	FLAG_VOICE	FLAG_ATV	0.071	0.283
7	FLAG_INET	FLAG_ATV	0.210	0.369
8	FLAG_INET, FLAG_VOICE	FLAG_ATV	0.055	0.382
9	FLAG_VOICE	FLAG_INET	0.145	0.576
10	FLAG_ATV	FLAG_INET	0.210	0.755
11	FLAG_ATV, FLAG_VOICE	FLAG_INET	0.055	0.778

Obr. 15 Výstup z operátoru Create Association Rules pro databázi Královehradeckého kraje

Jako poslední způsob určení predikce zákazníků v závislosti na informacích o preferencích spotřebitelů uvádím výstup z prostředí RapidMineru a operátoru Create Association Rules. Jak již bylo zmíněno, primární službou ohledně počtu zákazníků i dosahovaného zisku je služba FLAG_INET. Informace, které nám poskytuje obrázek číslo 14, doplňují pravdivostní hodnotu údajů, uvedených v tabulce číslo 6 ve sloupcích *procentuální zastoupení a pozice dle procent*. Zaměříme se pouze na asociační pravidla číslo 9, 10 a 11, které nám uvádějí pravděpodobnost pořízení služby FLAG_INET v závislosti na dříve zakoupených službách. Asociační pravidlo číslo 9 popisuje tuto situaci: v případě, že si zákazník pořídí nejprve službu FLAG_VOICE, existuje zde 57,6 % šance na následné zakoupení služby FLAG_INET. Asociační pravidlo číslo 10 popisuje situaci, kdy si zákazník nejdříve pořídí samostatnou službu FLAG_ATV. V tomto případě je šance na zakoupení služby FLAG_INET u stejného zákazníka 75,5 %. Nejpodstatnějším asociačním pravidlem je pro nás pravidlo číslo 11, které popisuje kombinaci dvou služeb a následné zakoupení FLAG_INET. V případě, že si zákazník zakoupí kombinaci služeb FLAG_ATV a FLAG_VOICE, je zde 77,8 % šance na pořízení služby FLAG_INET. Jak je ale dobře viditelné ve sloupci *Support* (obrázek číslo 14), procentuální zastoupení v databázi zákazníků Královehradeckého kraje je velmi nízké.

5.4.3 Získání zákazníků v Královehradeckém kraji

Vzhledem ke vztahům v databázi uživatelů Královehradeckého kraje, získaných pomocí metod data miningu, lze doporučit firmě UVXA určité způsoby, které by mohly vést k získání nových zákazníků v daném regionu. Základem pro získání nových zákazníků v regionu by měla být snaha vytvoření určitých benefitů, které by spotřebitele mohly vést k zakoupení služby. Po vytvoření benefitů, zacílených čistě na spotřebitele v Královehradeckém kraji, by bylo nutné, aby firma UVXA v daném regionu zvýšila úroveň prezentace svých produktů. Pokud má firma UVXA zájem expandovat do okresů Královehradecké kraje, ve kterých dosud nemá zastoupení (Jičín, Trutnov, Náchod a Rychnov nad Kněžnou), měla by prioritně nabízet službu FLAG_INET. Služba FLAG_INET by měla být nabízena buďto samostatně za zvýhodněnou cenu, nebo v určitých balíčcích složených z více služeb. Při předpokládání stejného ekonomického uvažování zákazníků v daném regionu, může firma UVXA vytvořit cenově zvýhodněný balíček služeb FLAG_ATV, FLAG_VOICE a FLAG_INET. Kombinaci služeb zacílených na potenciální zákazníky Královehradeckého kraje jsem zvolil na základě její procentuální oblíbenosti zakoupení u zákazníků z Hradce Králové. Tuto zvýhodněnou kombinaci služeb by následně mohla zasílat potenciálním zákazníkům v Královehradeckém kraji v tištěné či e-mailové podobě. Další možností by bylo vytvoření reklamy na dané produkty na e-shopu firmy UVXA. Při návštěvě e-shopu by potenciálnímu zákazníkovi z lokality Královehradeckého kraje byla jako první zobrazena speciální nabídka zvýhodněných služeb a jednoduchý způsob provedení objednávky. Základním předpokladem pro získání nových zákazníků by zajisté byla předběžná analýza veškeré konkurence v oblasti poskytování telekomunikačních služeb v regionu a následné stanovení nejnižší ceny. Při delším trvání cenové konkurence by mohla firma UVXA docílit vytlačení ostatních poskytovatelů telekomunikačních služeb a následné ovládnutí trhu v regionu.

5.5 Analýza Karlovarského kraje

V porovnání s Královehradeckým krajem je na tom nejzápadnější kraj České republiky o poznání lépe. Počet zákazníků firmy UVXA dosahuje v Karlovarském kraji počtu 5932, čímž podstatně překonává Královehradecký kraj, kde se počet uživatelů vyšplhal na 3604. Rozdíl mezi počty zákazníků krajů je na první počet patrný, a to samé platí i pro zisky plynoucích z jednotlivých krajů. Databáze Karlovarského kraje obsahuje 1,64 krát více zákazníků, kteří mají zakoupeno o 12 128 služeb více než zákazníci v Královehradeckém kraji. Zisky Karlovarského kraje pro firmu UVXA jsou o 1,68 krát vyšším než v kraji Královehradeckém. Firma UVXA má v Karlovarském kraji zastoupení ve dvou ze tří okresů (66,6 %), a to konkrétně v Karlových Varech a v Chebu. Jediným neobsazeným okresem je Sokolov, ve kterém nemá firma ani jednoho zákazníka. Pro analýzu Karlovarského kraje a následné určení nejlepší kombinace prodeje služeb použiji, stejně jako u předchozí analýzy, asociační pravidla a rozhodovací stromy. Na základě získaných informací budu hledat řešení, které by přimělo potencionální zákazníky v okrese Sokolov ke koupi produktů firmy UVXA.

Size	Support	Item 1	Item 2	Item 3	Item 4
1	0.947	FLAG_DTV			
1	0.576	FLAG_INET			
1	0.269	FLAG_ATV			
1	0.257	FLAG_VOICE			
2	0.546	FLAG_DTV	FLAG_INET		
2	0.255	FLAG_DTV	FLAG_ATV		
2	0.245	FLAG_DTV	FLAG_VOICE		
2	0.209	FLAG_INET	FLAG_ATV		
2	0.146	FLAG_INET	FLAG_VOICE		
2	0.069	FLAG_ATV	FLAG_VOICE		
3	0.198	FLAG_DTV	FLAG_INET	FLAG_ATV	
3	0.138	FLAG_DTV	FLAG_INET	FLAG_VOICE	
3	0.064	FLAG_DTV	FLAG_ATV	FLAG_VOICE	
3	0.053	FLAG_INET	FLAG_ATV	FLAG_VOICE	
4	0.050	FLAG_DTV	FLAG_INET	FLAG_ATV	FLAG_VOICE

Obr. 16 Výstup z operátoru FP-Growth pro databázi Karlovarského kraje

Obrázek číslo 16 zobrazuje výstup z operátoru FP-Growth, který nám udává hodnoty četností služeb pro Karlovarský kraj. Stejně tak jako u obrázku číslo 12, i zde první sloupec představuje kombinace služeb vyskytujících se v databázi jako zakoupené. Podstatným rozdílem oproti analýze minulého kraje je výskyt atributu FLAG_DTV, který je zapříčiněn zakoupením této služby u méně než 100 % uživatelů. I přes nezájem některých zákazníků o službu je FLAG_DTV nejoblíbenější volbou samostatné služby v Karlovarském kraji, a to

přesně v 94,7 % případech. Nejvíce preferovanou kombinací dvou služeb jsou produkty FLAG_DTV a FLAG_INET, které se u zákazníků vyskytují v 54,6 %. Z 19,8 % je v databázi firmy UVXA zastoupena trojkombinace služeb ve spojení FLAG_DTV, FLAG_INET, FLAG_ATV. Novinkou oproti Královehradeckému kraji je fakt, že v nejzápadnějším kraji České republiky existuje 5 % množina zákazníků, kteří mají zakoupené všechny služby najednou.

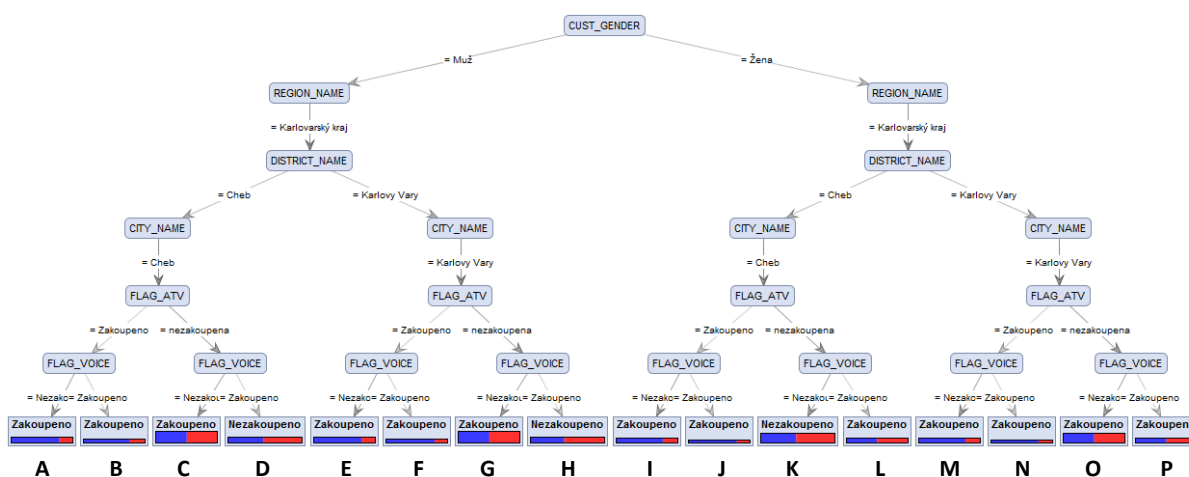
Size	Support	Item 1	Item 2	Item 3	Item 4	Cena	Zisk
1	0.947	FLAG_DTV				283	Zisk2
1	0.576	FLAG_INET				371	Zisk3
1	0.269	FLAG_ATV				143	Zisk 13
1	0.257	FLAG_VOICE				224	Zisk 15
2	0.546	FLAG_DTV	FLAG_INET			654	Zisk 1
2	0.255	FLAG_DTV	FLAG_ATV			426	Zisk 7
2	0.245	FLAG_DTV	FLAG_VOICE			507	Zisk 5
2	0.209	FLAG_INET	FLAG_ATV			514	Zisk 8
2	0.146	FLAG_INET	FLAG_VOICE			595	Zisk 9
2	0.069	FLAG_ATV	FLAG_VOICE			367	Zisk 14
3	0.198	FLAG_DTV	FLAG_INET	FLAG_ATV		797	Zisk 4
3	0.138	FLAG_DTV	FLAG_INET	FLAG_VOICE		878	Zisk 6
3	0.064	FLAG_DTV	FLAG_ATV	FLAG_VOICE		650	Zisk 11
3	0.053	FLAG_INET	FLAG_ATV	FLAG_VOICE		738	Zisk 12
4	0.050	FLAG_DTV	FLAG_INET	FLAG_ATV	FLAG_VOICE	1021	Zisk 10

Tab. 7 Četnosti služeb v Karlovarském kraji, jejich nacenění a ziskovost

Tabulka číslo 7 z převážné většiny vychází z výstupu operátoru FP_Growth. Mimo základních informací z výstupu operátoru FP_Growth je tabulka doplněna o informace o průměrných cenách za jednotlivé služby či za kombinace služeb. Průměrné hodnoty cen za služby se nachází ve sloupci označeném *cena*. Poslední sloupec tabulky úzce souvisí se sloupcem průměrných cen za služby, jelikož v závislosti na procentuálním zastoupení a průměrné ceně za služby určuje pořadí ziskovosti pro firmu UVXA. Ceny jsou uvedené v průměrných hodnotách ze stejného důvodu, jako u obsahově podobné kapitoly v analýze Královehradeckého kraje. Sloupec *Zisk* je opět uveden pouze ve slovní podobě a to ze stejného důvodu jako u sloupce *Zisk pro firmu UVXA* u tabulky číslo 5. Tabulka číslo 7 je oproti tabulce číslo 5 podstatně obsáhlejší, což zapříčinilo ne-stoprocentní zastoupení služby FLAG_DTV. Pro další analýzu máme v Karlovarské kraji na výběr z 12 různých kombinací služeb, přičemž každá z nich přináší firmě UVXA rozdílné příjmy. Nejvíce perspektivní je kombinace služeb FLAG_DTV a FLAG_INET, která je v databázi zastoupena z 54,6 % což je v porovnání s ostatními dvojkombinacemi služeb velmi markantní rozdíl.

Druhý největší zisk přináší firmě UVXA prodej samostatné služby FLAG_DTV, kterou má zakoupeno 94,7 % všech zákazníků Karlovarského kraje. Nejoblíbenější a zároveň nejziskovější trojkombinací tvoří služby FLAG_DTV, FLAG_INET a FLAG_ATV. Tato kombinace služeb zaujímá čtvrté místo na pomyslném žebříčku nejvíce ziskových kombinací a zároveň je v databázi Karlovarského kraje zastoupena z 19,8 %, což je na trojkombinaci služby velmi vysoké procento. V Karlovarském kraji se oproti Královehradeckému vyskytuje i kombinace všech čtyř hlavních nabízených služeb. Čtyřkombinaci má zakoupeno 5 % zákazníků a na žebříčku ziskovosti se nalézá na desáté pozici. Stejně tak, jako u Královehradeckého kraje, i v případě Karlovarského se zaměřím na případ, která přináší firmě UVXA nejvyšší zisky a to sice dvojkombinaci služeb FLAG_DTV a FLAG_INET. V následně použitém rozhodovacím stromu jsou za cílovou skupiny (label) označeny tyto dvě služby a jsou hledány jednotlivé zákonitosti, které vedou zákazníky ke koupi této služby.

5.5.1 Rozhodovací strom Karlovarský kraj



Obr. 17 Rozhodovací strom pro kombinaci služeb FLAG_DTV a FLAG_INET v Karlovarském kraji

Výše uvedený rozhodovací strom je zaměřen na zjištění zákonitostí, které vedou zákazníky k pořízení nejvíce ziskové kombinace služeb FLAG_DTV a FLAG_INET. Obrázek číslo 16 zobrazuje celý rozhodovací strom, jenž stejně tak, jako v případě Královehradeckého kraje, vnímá jako první atribut pohlaví zákazníka a tím pádem rozděluje databázi zákazníků firmy na muže/ženu. Rozhodovací strom pro Karlovarský kraj je podstatně obsáhlejší, než pro kraj Královehradecký a z tohoto důvodu jej lze nalézt v přílohách, kde je rozdělen do dvou částí, pro lepší vizualizaci.

Atribut DISTRICT_NAME je pro všechny muže i ženy společný a označuje kraj, ze kterého zákazníci pocházejí. První větvení do jednotlivých cílových skupin přináší atribut DISTRICT_NAME, kde se zákazníci dělí do okresů. Jelikož v každém okresu je zastoupeno pouze jedno město, označuje atribut CITY_NAME pouze název města, ze kterého zákazníci pocházejí. Dále již probíhá rozdělení zákazníků do konečné skupiny na základě dvou služeb, u kterých je zkoumáno zakoupení/nezakoupení dané služby. Ve výše uvedeném rozhodovacím stromu jsou pro rozhodovací kritérium zobrazeny pouze dvě ze čtyř nabízených služeb od firmy UVXA. Absence služeb FLAG_DTV a FLAG_INET je způsobena označením těchto dvou služeb jako label a tím pádem jsou hledány zákonitosti pro pořízení těchto dvou služeb jako celku. Stejně tak, jako u předchozí analýzy, i v tomto případě použijí pro detailnější popis tohoto rozhodovacího stromu následující tabulku, která popisuje přesné hodnoty u množin zákazníků, kteří si zakoupili kombinaci služeb FLAG_DTV a FLAG_INET.

Označení množiny	Počet nezakoupení	Počet zakoupení	Pozice dle zisku	Procentuální zakoupení	Pozice dle procent
Muži					
Množina A	71	256	4	78,29	2
Množina B	31	93	8	75,00	4
Množina C	444	466	1	51,21	6
Množina D	166	159	5	48,92	7
Množina E	76	273	3	78,22	3
Množina F	23	92	7	80,00	1
Množina G	430	454	2	51,36	5
Množina H	161	153	6	48,73	8
Ženy					
Množina I	59	191	4	76,40	4
Množina J	18	68	8	79,07	1
Množina K	370	349	2	48,54	8
Množina L	113	115	6	50,44	7
Množina M	61	201	3	76,72	3
Množina N	18	64	7	78,05	2
Množina O	345	360	1	51,06	5
Množina P	131	136	5	50,94	6

Tab. 8 Reprezentace číselných hodnot z rozhodovacího stromu pro Karlovarský kraj

V případě Karlovarského kraje a mužského pohlaví jsou nejpočetněji zastoupeny dvě množiny a to sice *množina C* a *množina G*. V případě *množiny C* je reprezentována následující cesta k cílové množině: (CUST_GENDER/MUŽ -> REGION_NAME/KARLOVARSKÝ KRAJ -> DISTRICT_NAME/CHEB -> CITY_NAME/CHEB -> FLAG_ATV/NEZAKOUPENO -> FLAG_VOICE/NEZAKOUPENO). Tato množina přináší firmě UVXA nejvyšší zisky u všech mužů z Karlovarského kraje a tím pádem i nejvyšší zisky v okresu Cheb. Pro množinu G, která

přináší nejvyšší zisky v okrese Karlovy Vary, je cesta k cílové množině následující:
(CUST_GENDER/MUŽ -> REGION_NAME/KARLOVARSKÝ KRAJ -> DISTRICT_NAME/Karlovy Vary -> CITY_NAME/Karlovy Vary -> FLAG_ATV/NEZAKOUPENO -> FLAG_VOICE/NEZAKOUPENO). Pokud porovnáme obě dvě cesty pro množinu C a množinu G, najdeme velmi výraznou spojitost. Zákazníci z obou okresů mají velmi podobné preference a k zakoupení dvojkombinace služeb FLAG_DTV a FLAG_INET přistupují povětšinou velmi stejnou cestou. Toto tvrzení lze dokázat na obrázku číslo 16, na kterém je na první pohled vidět, že grafické vyjádření konečných množin je pro oba okresy velmi podobné. Samozřejmostí zůstává, že hodnoty pořízení nejsou stejné, ale velice podobné.

U zákazníků ženského pohlaví je situace také velmi podobná. Hlavním rozdílem mezi zákazníky muž/žena je fakt, že i když mají ženy velmi podobné procentuální zakoupení služeb jako muži, služeb si pořizují výrazně méně. V případě žen jsou pak nejpočetněji zastoupeny dvě množiny a to sice *množina O* a *množina K*. Zajímavostí je, že i v tomto případě je cesta velmi podobná množinám C a G. Ve své podstatě bychom na základě tabulky číslo 8 a rozhodovacího stromu pro Karlovarský kraj na obrázku 16 mohli tvrdit, že existuje určitý „vzorec“ chování zákazníků, který jsme schopni detailně popsat. K detailnímu popisu vzorce nám slouží uvedené cesty k získání dvojkombinace služeb a také číselné hodnoty z tabulky. Pro firmu UVXA je tedy v případě mužů nejziskovější množina C a v případě žen množina O. Pokud bychom jako kritérium úspěšnosti množiny určili procentuální úspěšnost zakoupení dvojkombinace služeb, pak nejlepší množinou v případě mužů byla stanovena *množina F* a v případě žen množina J.

5.5.2 Asociační pravidla pro Karlovarský kraj

No.	Premises	Conclusion	Support	Confidence
21	FLAG_VOICE	FLAG_ATV	0.069	0.267
22	FLAG_DTV	FLAG_ATV	0.255	0.269
23	FLAG_INET	FLAG_DTV, FLAG_ATV	0.198	0.343
24	FLAG_INET, FLAG_VOICE	FLAG_DTV, FLAG_ATV	0.050	0.345
25	FLAG_INET	FLAG_ATV	0.209	0.362
26	FLAG_DTV, FLAG_INET	FLAG_ATV	0.198	0.362
27	FLAG_DTV, FLAG_INET, FLAG_VOICE	FLAG_ATV	0.050	0.364
28	FLAG_INET, FLAG_VOICE	FLAG_ATV	0.053	0.366
29	FLAG_VOICE	FLAG_DTV, FLAG_INET	0.138	0.536
30	FLAG_DTV, FLAG_VOICE	FLAG_INET	0.138	0.563
31	FLAG_VOICE	FLAG_INET	0.146	0.567
32	FLAG_DTV	FLAG_INET	0.546	0.576
33	FLAG_ATV, FLAG_VOICE	FLAG_DTV, FLAG_INET	0.050	0.732
34	FLAG_ATV	FLAG_DTV, FLAG_INET	0.198	0.735

Obr. 18 Výstup z operátoru Create Association Rules pro databázi Karlovarského kraje

Jako poslední způsob určení predikce zákazníků v Karlovarském kraji zvolím stejnou data miningovou metodu jako u analýzy Královehradeckého kraje. V závislosti na informacích o preferencích spotřebitelů uvádím výstup z prostředí RapidMineru a operátoru Create Association Rules. Jak již bylo zmíněno, primární službou ohledně počtu zákazníků i dosahovaného zisku je dvojkombinace služeb FLAG_DTV a FLAG_INET. Informace, které nám poskytuje obrázek číslo 17, opět doplňují pravdivostní hodnotu údajů, uvedených v tabulce číslo 8 ve sloupcích *procentuální zastoupení* a *pozice dle procent*. Výčet asociačních pravidel je v případě Karlovarského kraje velmi obsáhlý a z tohoto důvodu je uveden v příloze. V našem případě se zaměříme pouze na asociační pravidla číslo 29, 33 a 34, které nám uvádí pravděpodobnost pořízení dvojkombinace služeb FLAG_DTV a FLAG_INET.

Asociační pravidlo číslo 22 popisuje následující situaci: pokud si zákazník nejdříve zakoupí službu FLAG_VOICE, existuje zde 53,6 % šance na zakoupení dvojkombinace služeb FLAG_DTV a FLAG_INET. Asociační pravidlo číslo 33 popisuje situaci, kdy má zákazník zakoupené všechny čtyři služby naráz. Pokud při zakupování služeb jako první zvolí zákazník kombinaci služeb FLAG_ATV, FLAG_VOICE, pak je ze 73,2% možné, že si zakoupí také

dvojkombinaci služeb FLAG_DTV a FLAG_INET. Tento případ se v případě zákazníků Karlovarského kraje však vyskytuje velmi zřídka, a to přesně v 5 % případů. Posledním asociační pravidlo zaměřené na možnost zakoupení dvojkombinace služeb FLAG_DTV a FLAG_INET je asociační pravidlo 34. Pravidlo 34 nabývá nejvyšší hodnoty pravděpodobnosti, a to konkrétně v případě, že si nejprve zákazník zakoupí službu FLAG_ATV. V tomto případě lze očekávat, že si zákazník zakoupí i dvojkombinaci služeb FLAG_DTV a FLAG_INET, a to přesně na 73,5 %. Velmi podstatný je i fakt, že poslední asociační pravidlo se v databázi vyskytuje nejčastěji, a to přesně v 19,8 % případů.

5.5.3 Získání zákazníků v Královehradeckém kraji

Vzhledem k informacím o vztazích v databázi uživatelů Karlovarského kraje, které jsem získal pomocí metod data miningu, lze firmě UVXA doporučit určité způsoby, kterými by mohla docílit získání nových zákazníků v daném regionu. V případě Karlovarského kraje bych firmě UVXA doporučil zaměřit se na okres Sokolov a při marketingové kampani aplikovat znalosti ze zbývajících okresů Karlovarského kraje. Jak již bylo uvedeno výše, zákazníci z okresu Karlovy Vary a okresu Cheb mají velmi podobné preference při pořizování služeb od firmy UVXA, a proto lze předpokládat, že podobné ekonomické uvažování budou mít i potencionální zákazníci v okresu Sokolov.

Prioritně by, dle mého názoru, měla firma UVXA nabízet v okresu Sokolov dvojkombinaci služeb FLAG_DTV a FLAG_INET, která jí přináší nejvyšší zisk z prodeje kombinace služeb a zároveň je nejčastěji pořizovanou kombinací v Karlovarském kraji. Tato dvojkombinace služeb by měla být nabízena buďto samostatně za zvýhodněnou cenu, nebo v určitých balíčcích, složených z více služeb. Jak bylo zjištěno pomocí asociačních pravidel, příkladem balíčku více služeb by například mohl být případ posledního asociačního pravidla (pravidla 34). V tomto případě bych navrhol firmě UVXA vytvoření cenové akce na službu FLAG_ATV a její následné prezentování v okresu Sokolov. V případě koupě služby FLAG_ATV totiž vzniká firmě 73,5 % šance, že si zákazník zakoupí i kombinaci služeb FLAG_DTV a FLAG_INET, která jí přináší nejvyšší zisky. Zvýhodněnou kombinaci služeb FLAG_DTV a FLAG_INET by bylo možné společně s akcí na službu FLAG_ATV zasílat potencionálním zákazníkům v Karlovarském kraji v tištěné či e-mailové podobě. Stejně tak, jako v minulém

případě analýzy, by další možností by bylo vytvoření reklamy na dané produkty na e-shopu firmy UVXA. Při návštěvě e-shopu by potencionálnímu zákazníkovi z lokality Karlovarského kraje byla jako první zobrazena speciální nabídka zvýhodněných služeb a jednoduchý způsob provedení objednávky. V Karlovarském kraji by také bylo možné propagovat svou nabídku prostřednictvím regionální televize, rozhlasu a tisku. Vzhledem k osobním zkušenostem vím, že regionální sdělovací prostředky mají v Karlovarském kraji velký vliv a jsou u obyvatelů kraje velmi oblíbené. Základním předpokladem pro získání nových zákazníků by zajisté byla předběžná analýza veškeré konkurence v oblasti poskytování telekomunikačních služeb v regionu a následné stanovení nejnižší ceny. V případě okresu Sokolov by, dle mého názoru, bylo velmi složité získat nové zákazníky, a to z důvodu dominantního vlivu místního poskytovatele telekomunikačních služeb, který působí velkou konkurenci firmě UVXA i v okresech Cheb a Karlovy Vary.

6. Závěr

Získávání užitečných informací z databází pomocí data miningu je založeno na jednoduchém principu aplikování určitých operátorů, a to například v programu RapidMiner. Největší výhodou data miningu je možnost analyzování obsáhlých databází, u kterých by jinak nebylo možné potřebné užitečné informace získat. Většina softwarů, specializovaných na data mining, přímo v uživatelském prostředí velmi dobře vysvětluje fungování jednotlivých operátorů, tudíž pro začátečníka není nutné dlouhé studium teorie. Výstupy z jednotlivých operátorů či soustavy operátorů tvořících proces, přináší velmi užitečné informace. Na základě těchto informací je následně firma v konkurenčním prostředí na lepším postavení a je schopna na své potencionální zákazníky zacílit správné produkty. Vhodnou volbou produktů pak může dosahovat velmi vysoké úspěšnosti při získávání nových zákazníků a tím i zvyšovat svůj zisk. Veškeré užitečné vztahy a informace, získané z databáze, je však nutné patřičně prezentovat, k čemuž velmi výrazným způsobem napomáhá provozování vhodného e-shopu, správné pochopení fungování E-commerce, jejich jednotlivých druhů a způsobu prezentace nabídky produktů na internetu.

Výstupem z praktické části této bakalářské práce je analýza databáze firmy UVXA, která může posloužit firmě k získání nových zákazníků ve slabých okresech České republiky. Detailní rozbor byl uveden pro dva kraje, ve kterých dosahuje firma UVXA nejnižších zisků. Dle mého názoru, je postup při zjišťování důležitých informací z databáze jednotlivých krajů velmi dobře zpracován a přináší klíčové informace, z kterých může firma UVXA výrazným způsobem profitovat. Hlavní předností vypracované analýzy krajů je fakt, že při dodržení stejného postupu je možná aplikace i na ostatní kraje České republiky a tím pádem vytvoření celorepublikového průzkumu, obsahujícím kompletní informace o preferencích spotřebitelů.

Vzhledem k velkému objemu dat, které mi byly poskytnuty, bych chtěl tuto bakalářskou práci v navazujícím studiu rozšířit na práci diplomovou. Diplomová práce by byla zaměřena čistě na firmu UVXA a na vytvoření komplexní analýzy preferencí zákazníků firmy. Ta by nebyla zaměřena jen na Českou republiku, ale i na určité zahraniční státy, ve kterých má firma UVXA také velký počet zákazníků.

6 Seznam použitých zdrojů

BRAMER, Max. *Principles of Data Mining*. 2. London: Springer London Heidelberg New York Dordrecht, 2007. ISBN 978-1-4883-8.

FAYYAD, Usama. *Advances in knowledge discovery and data mining*. Menlo Park, California 94025: The MIT press, 1996. ISBN 972-0-262-56097-9.

ALPAYDIN, Ethem. *Introduction to Machine Learning*. 2. London, England: The MIT Press, 2010. ISBN 9780262315548.

BERKA, Petr. *Dobývání znalostí z databází*. 1. Praha: Academia, 2005. ISBN 80-200-1062-9.

KOMPRDOVÁ, Klára. *Rozhodovací stromy a lesy* [online]. 1. Brno: Brno: akademické nakladatelství CERM, 2012 [cit. 2017-04-02]. ISBN 978-80-7204-785-7. Dostupné z: <https://www.iba.muni.cz/res/file/ucebnice/komprdova-rozhodovaci-stromy-lesy.pdf>

LAUDON, Kenneth C. a Carol TRAVER. *E-commerce Essentials*. 9. England: Pearson, 2013. ISBN 0-273-77935-4.

SUCHÁNEK, Petr. *E-commerce: Elektronické podnikání a koncepce elektronického obchodování*. 1. Praha 4: Ekopress, 2012. ISBN 978-80-86929-84-2.

B2C. In: *Shopcentrik.cz* [online]. Praha, 2012 [cit. 2017-03-21]. Dostupné z: <http://www.shopcentrik.cz/slovník/b2c.aspx>

Zlatý věk e-shopů v Česku skončil. In: *Tyden.cz* [online]. Praha: tyden.cz/Kamila Jušková, 2016 [cit. 2017-03-17]. Dostupné z: http://www.tyden.cz/rubriky/byznys/cesko/zlaty-vek-e-shopu-v-cesku-skoncil_384822.html

B2B: Business to business. In: *Shopcentrik.cz: www.shopcentrik.cz/slovník* [online]. Karolinská 661/4 186 00 Praha 8 - Karlín, 2012 [cit. 2017-03-16]. Dostupné z: <http://www.shopcentrik.cz/slovník/b2b.aspx>

Jak funguje a co přináší B2B řešení: Interval - Jak funguje a co přináší B2B řešení. *Interval.cz* [online]. Interval: Radek Šilhavý, 2002 [cit. 2017-03-11]. Dostupné z: <https://www.interval.cz/clanky/jak-funguje-a-co-prinasi-b2b-reseni/>

7 Seznam obrázků

Obrázek 1 - Technologický postup výše uvedeného dobývání znalostí z databází	17
Obrázek 2 - Manažerský postup výše uvedeného dobývání znalostí z databází	18
Obrázek 3 - Grafické znázornění vlastního rozhodovacího stromu podle kritérií v tabulce	24
Obrázek 4 - Grafické znázornění rozhodovacího stromu v programu RapidMiner	25
Obrázek 5 - Znázornění modelu na určení asociačních pravidel v programu RapidMiner	28
Obrázek 6 - výstup operátoru <i>FP-Growth (Frequent Pattern Growth)</i>	29
Obrázek 7 - výstup operátoru Create Association Rules	30
Obrázek 8 - Aplikace aglomerativního shlukování v programu RapidMiner	32
Obrázek 9 - Grafický výstup z modelu pro určení aglomerativního shlukování	33
Obrázek 10 - Grafické znázornění počtu uživatelů na mapě České republiky	36
Obrázek 11 - Grafické znázornění zisků firmy UVXA v krajích na mapě České republiky	37
Obrázek 12 Grafické propojení počtu uživatelů v okresech a zisků v krajích	49
Obrázek 13 Výstup z operátoru FP-Growth pro databázi Královehradeckého kraje	40
Obrázek 14 Rozhodovací strom pro službu FLAG_INET v Královehradeckém kraji	42
Obrázek 15 Výstup z operátoru Create Association Rules pro databázi Královehradeckého kraje	44
Obrázek 16 Výstup z operátoru FP-Growth pro databázi Karlovarského kraje	46
Obr. 17 Rozhodovací strom pro kombinaci služeb FLAG_DTV a FLAG_INET v Karlovarském kraji	48
Obrázek 18 Výstup z operátoru Create Association Rules pro databázi Karlovarského kraje	51

8 Seznam tabulek

Tabulka 1 - Zadaná data pro ukázky rozhodovacího stromu	23
Tabulka 2 - Zadaná data pro ukázkou fungování asociačních pravidel	27
Tabulka 3 - Ukázková data pro aglomerativní shlukování	32
Tabulka 4 Popis atributů databáze firmy UVXA	34
Tabulka 5 - Četnosti služeb v Královehradeckém kraji a jejich nacenění	41
Tabulka 6 - Reprezentace číselných hodnot z rozhodovacího stromu Královehradecký kraj .	46
Tabulka 7 - Četnosti služeb v Karlovarském kraji, jejich nacenění a ziskovost	47
Tabulka 8 - Reprezentace číselných hodnot z rozhodovacího stromu pro Karlovarský kraj	49

9 Seznam příloh

Příloha 1 – Rozhodovací strom Karlovarský kraj	59
Příloha 2 – popis rozhodovacího stromu Karlovarského kraje v prostředí RapidMiner	60
Příloha 3 – rozhodovací strom pro Karlovarský kraj	61

10 Přílohy

The screenshot displays the RapidMiner Studio interface with a workflow in the Design view. The workflow consists of the following operators:

- Read Excel**: Reads data from an Excel file.
- Replace Missing Val.**: Replaces missing values in the data.
- Set Role**: Sets the role of attributes.
- ID3**: Performs decision tree classification.

The interface includes several panels:

- Repository**: Shows data sources like Samples, DB, Local Repository, and Cloud Repository.
- Operators**: A search-based list of operators, currently showing 'Set Role' and various Selection, Generation, Examples, Sampling, and Table operators.
- Parameters**: A configuration panel for the 'Process' operator, including settings for log verbosity, logfile, resultfile, random seed, send mail, and encoding.
- Help**: A panel providing a synopsis and description of the 'Process' operator, stating it is the root operator of every process.

At the bottom of the interface, there is a message: "Leverage the Wisdom of Crowds to get operator recommendations based on your process design!" with an "Activate Wisdom of Crowds" button.

Příloha 1 – Rozhodovací strom Karlovarský kraj (zdroj: Autor)

File Edit Process View Connections Cloud Settings Extensions

Views: Design Results

Result History ExampleSet (Read Excel) Tree (ID3)

Graph

Description

Annotations

Tree

```

CUST_GENDER = Muž
| REGION_NAME = Karlovarský kraj
| | DISTRICT_NAME = Cheb
| | | CITY_NAME = Cheb
| | | | FLAG_ATV = Zakoupeno
| | | | | FLAG_VOICE = Nezakoupeno: Zakoupeno (Zakoupeno=256, Nezakoupeno=71)
| | | | | FLAG_VOICE = Zakoupeno: Zakoupeno (Zakoupeno=93, Nezakoupeno=31)
| | | | | FLAG_ATV = nezakoupena
| | | | | | FLAG_VOICE = Nezakoupeno: Zakoupeno (Zakoupeno=466, Nezakoupeno=444)
| | | | | | FLAG_VOICE = Zakoupeno: Nezakoupeno (Zakoupeno=159, Nezakoupeno=166)
| | | DISTRICT_NAME = Karlovy Vary
| | | | CITY_NAME = Karlovy Vary
| | | | | FLAG_ATV = Zakoupeno
| | | | | | FLAG_VOICE = Nezakoupeno: Zakoupeno (Zakoupeno=273, Nezakoupeno=76)
| | | | | | FLAG_VOICE = Zakoupeno: Zakoupeno (Zakoupeno=92, Nezakoupeno=23)
| | | | | FLAG_ATV = nezakoupena
| | | | | | FLAG_VOICE = Nezakoupeno: Zakoupeno (Zakoupeno=454, Nezakoupeno=430)
| | | | | | FLAG_VOICE = Zakoupeno: Nezakoupeno (Zakoupeno=138, Nezakoupeno=161)
CUST_GENDER = Žena
| REGION_NAME = Karlovarský kraj
| | DISTRICT_NAME = Cheb
| | | CITY_NAME = Cheb
| | | | FLAG_ATV = Zakoupeno
| | | | | FLAG_VOICE = Nezakoupeno: Zakoupeno (Zakoupeno=191, Nezakoupeno=59)
| | | | | FLAG_VOICE = Zakoupeno: Zakoupeno (Zakoupeno=68, Nezakoupeno=18)
| | | | | FLAG_ATV = nezakoupena
| | | | | | FLAG_VOICE = Nezakoupeno: Nezakoupeno (Zakoupeno=349, Nezakoupeno=370)
| | | | | | FLAG_VOICE = Zakoupeno: Zakoupeno (Zakoupeno=115, Nezakoupeno=113)
| | | DISTRICT_NAME = Karlovy Vary
| | | | CITY_NAME = Karlovy Vary
| | | | | FLAG_ATV = Zakoupeno
| | | | | | FLAG_VOICE = Nezakoupeno: Zakoupeno (Zakoupeno=201, Nezakoupeno=61)
| | | | | | FLAG_VOICE = Zakoupeno: Zakoupeno (Zakoupeno=64, Nezakoupeno=18)
| | | | | FLAG_ATV = nezakoupena
| | | | | | FLAG_VOICE = Nezakoupeno: Zakoupeno (Zakoupeno=360, Nezakoupeno=345)
| | | | | | FLAG_VOICE = Zakoupeno: Zakoupeno (Zakoupeno=136, Nezakoupeno=131)

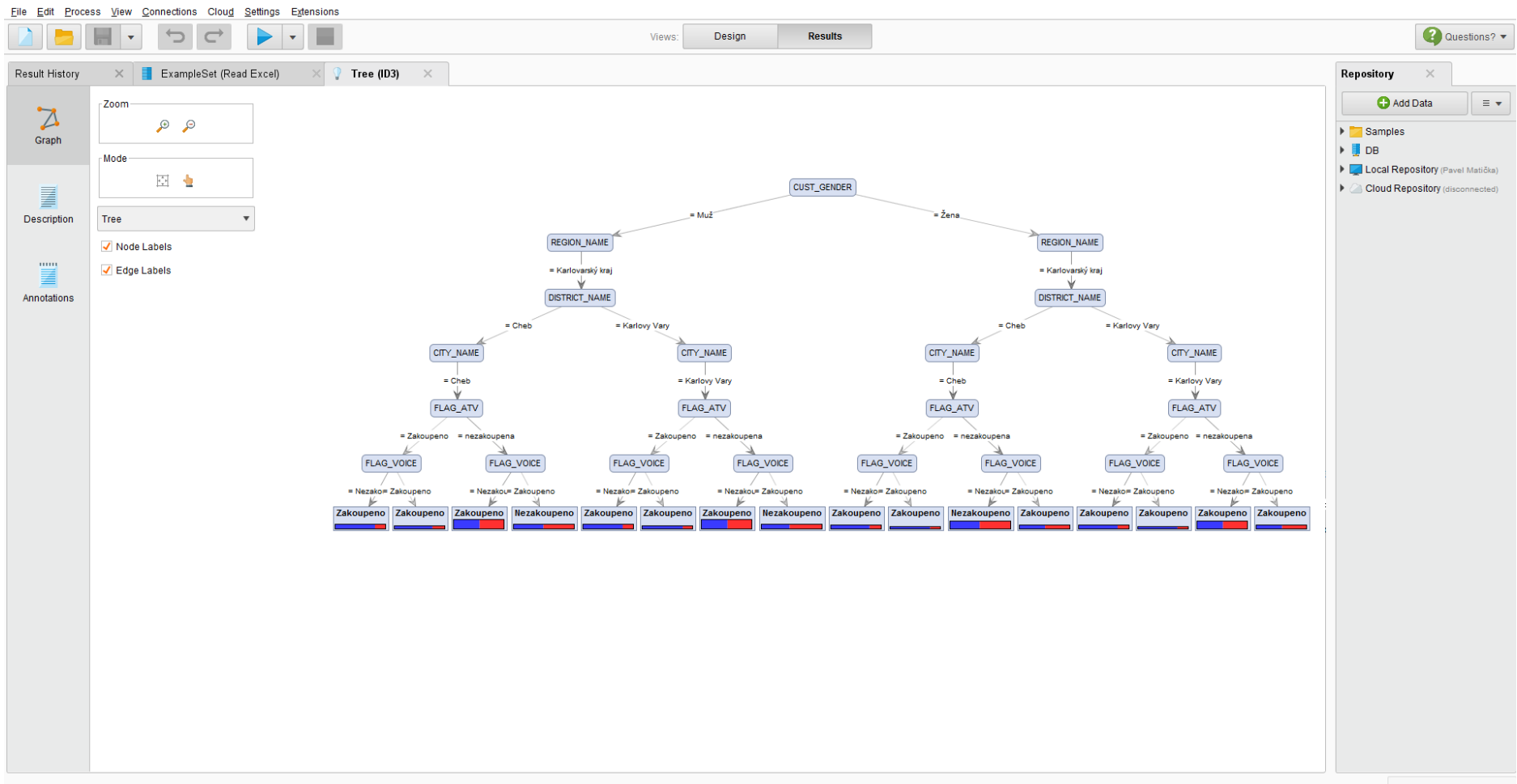
```

Repository

Add Data

- Samples
- DB
- Local Repository (Pavel Matička)
- Cloud Repository (disconnected)

Příloha 2 – popis rozhodovacího stromu Karlovarského kraje v prostředí RapidMiner (zdroj: Autor)



Příloha 3 – rozhodovací strom pro Karlovarský kraj (zdroj: Autor)