

UNIVERZITA PARDUBICE
FAKULTA EKONOMICKO-SPRÁVNÍ

Klasifikace studentů Fakulty ekonomicko-správní

Bc. Jana Čechová

DIPLOMOVÁ PRÁCE

2009

Univerzita Pardubice
Fakulta ekonomicko-správní
Ústav systémového inženýrství a informatiky
Akademický rok: 2008/2009

ZADÁNÍ DIPLOMOVÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Jana ČECHOVÁ**
Studijní program: **N6209 Systémové inženýrství a informatika**
Studijní obor: **Regionální a informační management - Regionální management**
Název tématu: **Klasifikace studentů Fakulty ekonomicko-správní**

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

- a) Popis současného stavu (možnosti studia na Fakultě ekonomicko - správní (FES) - studijní obory a programy, základní statistiky);
- b) návrh faktorů, které charakterizují studenty FES;
- c) sběr dat, vyhodnocení získaných dat a návrh datové matice;
- d) klasifikace studentů FES (charakteristika vybrané metody, návrh modelu a jeho verifikace ve zvoleném programovém prostředí);
- e) analýza výsledků.

Rozsah grafických prací:


Rozsah pracovní zprávy:

Forma zpracování diplomové práce: **tištěná/elektronická**

Seznam odborné literatury:

- 1) BERKA, P. Dobývání znalostí z databází. 1. vyd. Praha: Academia, 2003. ISBN 80-200-1062-9.
- 2) HAN, J., KAMBER, M. Data Mining: Concepts and Techniques. 2. vyd. San Francisco: Morgan Kaufmann, 2006. ISBN 1-55860-901-6.
- 3) MELOUN, M. - MILITKÝ, J. Kompendium statistického zpracování dat: metody a řešené úlohy včetně CD. 1. vyd. Praha: Academia Praha, 2002. ISBN 80-200-1008-4.
- 4) O'SULLIVAN, E., RASSEL, G., R., BERNER, M. Research Methods for Public Administration. 5. vyd. New York: Pearson Education, 2008. ISBN 978-0-321-43137-0.
- 5) RUD, O. P. Data Mining: Praktický průvodce dolováním dat pro efektivní prodej, cílený marketing a podporu zákazníků (CRM). 1. vyd. Praha: Computer Press, 2001. ISBN 80-7226-577-6.
- 6) ŘEZÁNKOVÁ, H. Analýza dat z dotazníkových šetření. 1. vyd. Praha: Professional Publishing 2007. ISBN 978-80-86946-49-8.

Vedoucí diplomové práce:



Ing. Miloslava Kašparová, Ph.D.
Ústav systémového inženýrství a informatiky

Datum zadání diplomové práce:


6. října 2008

Termín odevzdání diplomové práce:

1. května 2009


doc. Ing. Renáta Myšková, Ph.D.
děkanka

L.S.


doc. Ing. Jiří Křupka, Ph.D.
vedoucí ústavu

V Pardubicích dne 6. října 2008

Prohlašuji:

Tuto práci jsem vypracovala samostatně. Veškeré literární prameny a informace, které jsem v práci využila, jsou uvedeny v seznamu použité literatury.

Byla jsem seznámena s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorský zákon, zejména se skutečností, že Univerzita Pardubice má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1. autorského zákona, a s tím, že pokud dojde k užití této práce mnou nebo bude poskytnuta licence o užití jinému subjektu, je Univerzita Pardubice oprávněna ode mne požadovat přiměřený příspěvek na úhradu nákladů, které na vytvoření díla vynaložila, a to podle okolností až do jejich skutečné výše.

Souhlasím s prezenčním zpřístupněním své práce v Univerzitní knihovně Univerzity Pardubice.

V Pardubicích dne 30. 4. 2009

Jana ČECHOVÁ

Tímto bych ráda poděkovala vedoucí své diplomové práce Ing. Miloslavě Kašparové, Ph.D. za cenné rady a praktické připomínky při vytváření této diplomové práce.

Můj dík patří i mé rodině a přátelům, kteří mě v průběhu studia neustále všestranně podporují.

Jana ČECHOVÁ

Abstrakt

Tato diplomová práce se zabývá klasifikací studentů FES. Studenti jsou nejdříve osloveni dotazníkem, pomocí kterého jsou nasbírána potřebná data pro dotazníkové šetření. Odpovědi studentů pak tvoří datové matice, na jejichž základě je prováděno modelování závislostí v prostředí SPSS Clementine 10.1. Použita je shluková analýza a rozhodovací stromy. Na závěr jsou zhodnoceny dosažené výsledky práce.

Klíčová slova

Klasifikace; rozhodovací stromy; shluková analýza; dotazník

Title

Classification of Students Faculty of Economics and Administration

Abstract

This thesis deals with the classification of students FES. Students are first approached with the questionnaire, which collected the necessary data for the questionnaire survey. Replies students will form a data matrix on the basis of dependency modeling is carried out in SPSS Clementine 10.1. Decision trees and cluster analysis are used. In conclusion, the results obtained are assessed work.

Keywords

Classification, decision trees, cluster analysis, questionnaire

Obsah

ÚVOD.....	9
1. UNIVERZITA PARDUBICE	10
1.1 FAKULTA EKONOMICKO-SPRÁVNÍ.....	10
1.2 VÝVOJ POČTU STUDENTŮ.....	13
2. STRUKTUROVÁNÍ CÍLE PRÁCE.....	14
3. ZÍSKÁVÁNÍ DAT	15
3.1 DOTAZOVÁNÍ.....	15
3.2 DOTAZOVACÍ TECHNIKY	16
3.3 KLASIFIKACE DAT.....	18
3.3.1 Škály měření a typy proměnných.....	19
3.4 NÁVRH FAKTORŮ CHARAKTERIZUJÍCÍCH STUDENTY	20
3.4.1 Osobní dotazování.....	20
3.4.2 Studijní program.....	20
3.4.3 Hodnocení spokojenosti.....	21
4. VYHODNOCENÍ DOTAZNÍKOVÉHO ŠETŘENÍ.....	22
4.1 ÚSPĚŠNOST SBĚRU DAT.....	22
4.2 DATOVÁ MATICE.....	23
4.2.1 Rozvržení datového souboru <i>Osobni_dotazovani</i>	23
4.2.2 Rozvržení datového souboru <i>Studijni_program</i>	24
4.2.3 Rozvržení datového souboru <i>Hodnoceni_spokojenosti</i>	25
4.3 ANALÝZA PROMĚNNÝCH.....	26
4.3.1 Absolutní četnost.....	26
4.3.2 Prvotní analýza.....	27
5. NÁVRH KLASIFIKAČNÍHO MODELU A POPIS METOD.....	36
5.1 OBECNÉ DOBÝVÁNÍ ZNALOSTÍ.....	36
5.2 SCHÉMA MODELU	37
5.3 METODY KLASIFIKACE.....	37
5.3.1 Rozhodovací stromy.....	38
5.3.2 Shluková analýza	44
6. KLASIFIKACE STUDENTŮ.....	46
6.1 NAČTENÍ DAT.....	46
6.1.1 Nahrazení chybějících hodnot	48
6.2 ROZHODOVACÍ STROMY.....	48

6.2.1	<i>Osobní dotazování</i>	49
6.2.1.1	<i>Spokojenost s cenou kolejí</i>	49
6.2.1.2	<i>Spokojenost s cenou jídla</i>	51
6.2.1.3	<i>Financování</i>	57
6.2.2	<i>Studijní program</i>	58
6.2.2.1	<i>Opakování předmětů</i>	59
6.2.2.2	<i>Opakování matematických předmětů</i>	61
6.2.2.3	<i>Opakování infromatických předmětů</i>	63
6.2.2.4	<i>Studium v zahraničí</i>	64
6.2.2.5	<i>Vytyčené cíle</i>	65
6.3	SHLUKOVÁ ANALÝZA	69
ZÁVĚR		71
SEZNAM LITERATURY A POUŽITÝCH ZDROJŮ		73
LITERATURA		73
OSTATNÍ ZDROJE		74
SEZNAM TABULEK		75
SEZNAM OBRÁZKŮ		76
SEZNAM ZKRATEK		78
SEZNAM PŘÍLOH		79
PŘÍLOHY		80

Úvod

Dosáhnout vysokoškolské vzdělání je v dnešní době jedním z kroků pro dobré pracovní uplatnění a budování kvalitní pracovní pozice na trhu. Je na každém studentovi jak naloží s příležitostí studovat na vysoké škole. Najdou se studenti, kteří se intenzivně věnují studiu a na ostatní aktivity nemají čas. Na druhou stranu existují i tací, kteří dokáží skloubit studium s prací či jinými užitečnými aktivitami.

Cílem této práce je klasifikovat studenty Fakulty ekonomicko-správní Univerzity Pardubice. Klasifikovat je nejen na úrovni studia, ale i v oblastech ostatních aktivit, které ke studentům patří. Sběr dat je tedy zaměřen na získání všeobecných informací o studentech. Realizován je dotazníkovým šetřením na přednáškách. Tím bude dosaženo velké návratnosti a zároveň bude ošetřeno, aby skladba vrácených dotazníků korespondovala se skladbou základního souboru.

Data získaná z dotazníkového šetření budou v podobě datových matic analyzována pomocí rozhodovacích stromů a shlukové analýzy. Jako vstupy pro tyto modely budou voleny vhodné atributy vztahující se k cílovým proměnným. Výstupem modelů bude pak klasifikace studentů dle cílových proměnných a třídy, do kterých budou studenti zařazeni. Veškeré modelování bude provedeno v prostředí SPSS Clementine 10.1.

Práce je rozdělena do sedmi kapitol. První kapitola pojednává o Univerzitě Pardubice a podává základní přehled o Fakultě ekonomicko-správní. Druhá kapitola vysvětluje cíl práce. Ve třetí kapitole je popsáno získávání dat od studentů. Jakou formou a kde jsou studenti osloveni. Ve čtvrté kapitole je již první zmínka o vyhodnocení dotazníkového šetření. V páté kapitole je teoreticky popsáno jakým způsobem bude probíhat modelování na získaných datech. Jsou zde popsány modely, kterými budou znalosti z dat generovány. Šestá kapitola s názvem klasifikace studentů popisuje již konkrétní postup při dolování dat. Jsou zde vytyčeny cílové proměnné, stěžejní pro modelování a popsány výstupy vytvořených modelů. Závěrečná kapitola hodnotí a krátce popisuje dosažené výsledky celkové klasifikace.

1. Univerzita Pardubice

Univerzita Pardubice rozvíjí téměř šedesátiletou tradici vysokého školství ve městě. Z jednofakultní školy chemického zaměření se vznikem nových fakult po roce 1990 stala instituce poskytující vysokoškolské vzdělání univerzitního typu. Od roku 1994 nese současný název Univerzita Pardubice [15].

1.1 Fakulta ekonomicko-správní

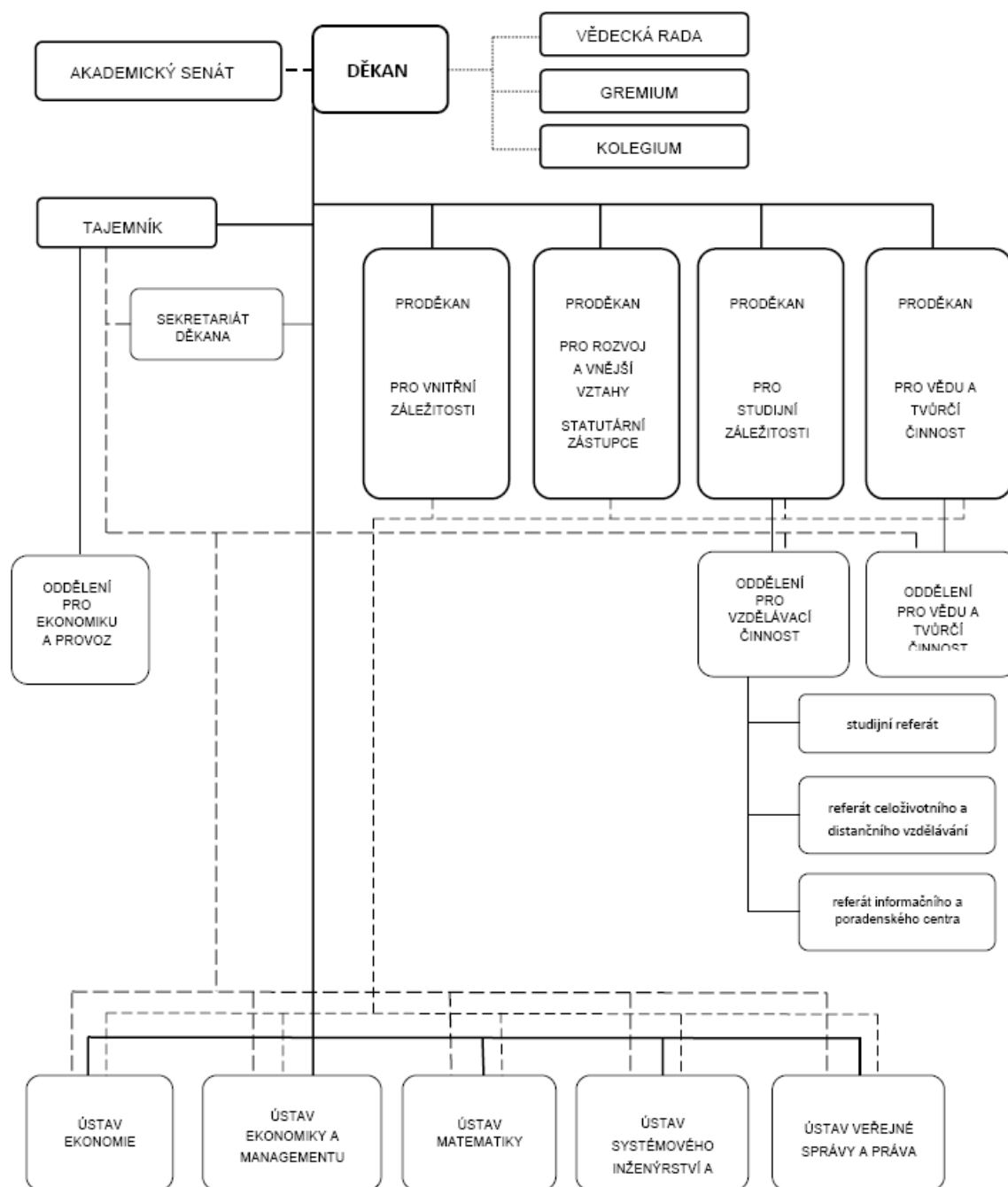
Fakulta prošla za tuto poměrně krátkou dobu velkým vývojem. Před 15 lety začalo na fakultě - tehdy Fakultě územní správy - studovat v jednom studijním oboru 50 studentů, v letošním roce studuje ve 3 studijních programech a 8 studijních oborech téměř 2 500 studentů. Je potěšující, že o studium na Fakultě ekonomicko-správní (FES) je mezi uchazeči ze středních škol stále velký zájem. V letošním roce si podalo přihlášku ke studiu v prezenční nebo kombinované formě přes 3 000 studentů. Fakultu za patnáct let existence opustilo 3 296 kvalifikovaných absolventů s bakalářským titulem nebo titulem „inženýr“ po absolvování magisterského studia. Fakulta rovněž nabízí těm nejzdatnějším studentům tři programy v doktorském studiu, nejvyšším stupni vysokoškolského vzdělávání [15].

Na Fakultě ekonomicko-správní pokračuje rozvoj v souladu s Dlouhodobým záměrem vzdělávací, vědecké, výzkumné, vývojové, umělecké a další tvůrčí činnosti Fakulty ekonomicko-správní Univerzity Pardubice do roku 2010. V současné době je Fakulta ekonomicko-správní co do počtu studentů (2 486) největší fakultou Univerzity Pardubice. Nastartovaný dynamický kvantitativní rozvoj z minulých let dosáhl v roce 2007 již žádoucího stavu a činnost fakulty se orientovala na zvyšování kvalitativní stránky jak ve vzdělávací tak vědecko-výzkumné oblasti.

Fakulta ekonomicko-správní nabízí tři bakalářské studijní programy a v jejich rámci devět studijních oborů, tři navazující magisterské studijní programy se šesti studijními obory a tři doktorské studijní programy (z toho dva v anglickém jazyce). Fakulta tak rozšířila svoji nabídku vzdělávání odborníků ve veřejné správě a systémovém inženýrství také o možnost uplatnění absolventů v podnikové praxi.

Fakulta v roce 2007 úspěšně získala akreditaci nových bakalářských studijních oborů studijního programu Ekonomika a management – obor Ekonomika a provoz podniku a obor Management ochrany podniku a společnosti (jak v prezenční tak kombinované formě), a rovněž akreditaci navazujícího magisterského studijního oboru Ekonomika a management podniku v programu Ekonomika a management. V rámci studijního programu Hospodářská politika a správa nově akreditovala navazující magisterský studijní obor Regionální rozvoj v prezenční formě studia. Dále byla v roce 2007 fakultě prodloužena doba platnosti akreditace bakalářského studijního programu Systémové inženýrství a informatika obor Informatika ve veřejné správě (v prezenční a kombinované formě studia), a také navazujícího magisterského studijního programu Systémové inženýrství a informatika obor Informatika ve veřejné správě (v kombinované formě). Fakulta také úspěšně obhájila reakreditaci doktorského studijního programu Hospodářská politika a správa v oboru Regionální a veřejná ekonomie (v prezenční a kombinované formě studia) jak v českém, tak anglickém jazyce (Regionaland Public Administration) [17].

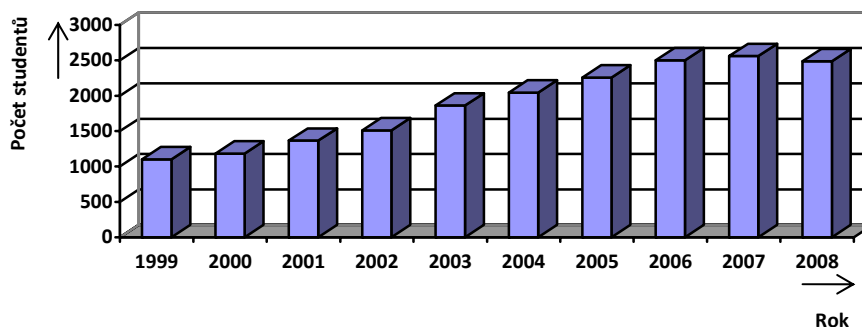
Níže uvedený obrázek (Obrázek 1) ukazuje základní strukturu Fakulty ekonomicko-správní. Podrobný přehled veškerých studijních programů (oborů) je uveden v Příloze č. 1.



Obrázek 1 – Organizační struktura FES [17]

1.2 Vývoj počtu studentů

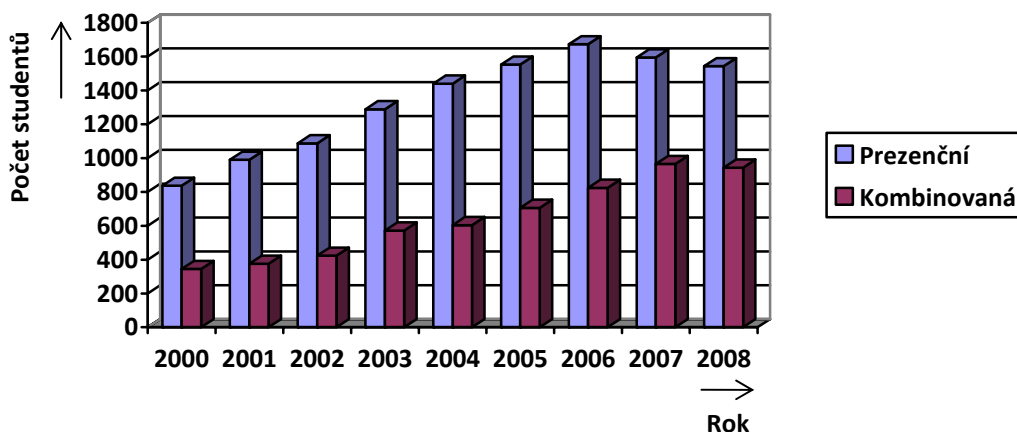
Celkové počty studentů FES Univerzity Pardubice v rozsahu let 1999 až 2008, vždy k 31. říjnu příslušného roku, je dokumentován následujícím Obrázkem 2.



Obrázek 2 - Vývoj počtu studentů [vlastní]

Jak je vidět, vývoj počtu studentů od roku 1999 neustále stoupal. Až v roce 2008 dochází k mírnému poklesu počtu studentů, který je způsoben kvalitativním zaměřením Univerzity Pardubice ve vzdělání a vědecko-výzkumné oblasti poté, co bylo dosaženo žádoucího stavu studentů v roce 2007.

Následující Obrázek 3 dokumentuje počty studentů FES kombinované a prezenční formy studia v letech 2000 až 2008.



Obrázek 3 – Vývoj počtu studentů kombinovaného a prezenčního studia [vlastní]

2. Strukturování cíle práce

Stěžejním cílem diplomové práce je klasifikace studentů na základě dat získaných dotazníkovým šetřením. Dotazníkové šetření bylo provedeno mezi studenty prvního až pátého ročníku za účelem zjištění zásadních informací ohledně studia, osobních přístupů studentů a hodnocení spokojenosti (studijních podmínek) na FES.

Jednotlivé otázky jsou koncipovány do třech dílů, tedy dílčích částí šetření:

- osobní dotazování,
- studijní program,
- hodnocení spokojenosti.

Záměrem je klasifikovat studenty na základě dílčích cílových proměnných a nakonec uspořádat a rozčlenit názory a vlastnosti zjištěné od studentů do tříd. Veškerá klasifikace je realizována v programu SPSS Clementine 10.1 pomocí technik a metod, které tento program umožňuje. Vše je realizováno na odpovědích studentů plynoucích pouze z první a druhé části dotazníkového šetření. Konkrétní výsledky dotazníkového šetření budou sloužit jako ukazatel determinující situaci a stanoviska studentů FES.

3. Získávání dat

3.1 Dotazování

Dotazování je v dnešní době nejčastěji používaná metoda sběru primárních dat ve výzkumech. Při aplikaci jednotlivých technik dotazování se používá dotazník.

Dotazník je formulář určený k pokud možno přesnému a úplnému zaznamenávání zjišťovaných informací. Dobrý dotazník je mnohem více než jenom soubor otázek, neboť usměrňuje proces dotazování. Při tvorbě dotazníku je účelné postupovat systematicky v postupných krocích. V praxi není nutné sekvenci kroků přesně dodržovat. Jejich vzájemná závislost totiž vede k tomu, že rozhodnutí přijatá v jednom určitém kroku mohou zpětně ovlivnit přístup v ostatních krocích. Kroky jsou následující [12]:

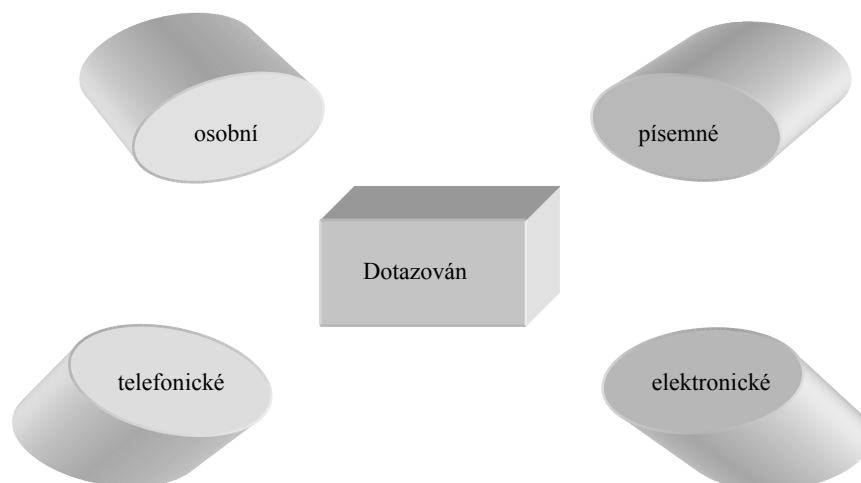
- stanovení údajů, které mají být zjištěny,
- stanovení procesu dotazování,
- hodnocení obsahu otázek,
- stanovení typu otázek,
- formulace otázek,
- stanovení struktury dotazníku,
- formální úprava dotazníku,
- pretest, korekce, konečný koncept.

Při tvorbě dotazníku je zásadní zaměřit se na sledování cíle práce. Se záměrem klasifikovat studenty FES byly strukturovány otázky na zjištění osobních údajů, studijních záležitostí a nakonec hodnocení spokojenosti. Proces dotazování byl stanoven jasně – oslovit studenty na přednáškách. Touto formou byla zajištěna maximální návratnost dotazníků. Po vytvoření otázek - základního konceptu dotazníku bylo vše konzultováno a proběhly případné úpravy. Dotazník je zaměřen především na uzavřené otázky. Uzavřené otázky nabízejí několik možných variant odpovědí, ze kterých si dotazovaný vybírá jednu nebo více odpovědí, které se nejvíce blíží jeho názoru [14]. Otázky otevřené, kde se respondent může vyjádřit svými vlastními slovy, dle vlastního uvážení, jsou v dotazníku použity minimálně. Struktura dotazníku je koncipována do tří částí - Část A (Osobní dotazování) viz. Příloha č. 2, Část B (Studijní program) viz. Příloha č. 3 a Část C (Hodnocení spokojenosti) viz. Příloha č. 4.

3.2 Dotazovací techniky

Techniky dotazníkového šetření je možné uskutečňovat čtyřmi hlavními typy technik šetření: osobní, písemné, telefonické a elektronické (Obrázek 4).

Výběr vhodného typu dotazování závisí na různých faktorech, především na charakteru a rozsahu zjišťovaných informací, skupině respondentů, časových a finančních limitech, kvalifikaci tazatele, atd. V praxi se většinou jednotlivé typy navzájem kombinují [3].



Obrázek 4 – Typy dotazování [3]

Pro svoji práci jsem zvolila techniku písemného dotazování. Písemné dotazování probíhalo na přednáškách a seminářích jednotlivých ročníků. Po představení a seznámení studentů s cílem dotazníkového šetření, byl studentům vysvětlen postup pro vyplňování dotazníku. Tento dotazník byl zodpovězen na místě. Vše bylo realizováno anonymně, studenti byli evidováni pouze při zpracovávání dotazníku formou pořadového čísla jako jednotlivé případy dotazníkového šetření.

Výhody písemného dotazování [12], [14]:

- široká distribuce, výběr může být uskutečněn ze souboru o širokém územním rozložení,
- nepřichází v úvahu možnost ovlivnění dotazovaného tazatelem,
- náklady jsou v porovnání s ostatními technikami šetření relativně nižší,
- oproti ostatním technikám vyžaduje relativně méně organizačních příprav,
- dotazovaný může věnovat zodpovězení otázek dostatek času a péče,
- při šetření konečných spotřebitelů jsou tímto způsobem některé vrstvy obyvatelstva snadněji dosažitelné, resp. lze jím kontaktovat osoby jinak nedosažitelné.

Nevýhody písemného dotazování [12], [14]:

- šetření probíhá zpravidla delší dobu (je třeba čekat na vrácení potřebného množství dotazníků),
- otázky musí být velmi jednoduché, snadno zodpověditelné, dotazník musí být relativně krátký, možnosti formulace jsou tedy značně omezeny,
- nelze získat údaje, které umožňuje přímé pozorování šetřeného subjektu,
- nelze kontrolovat, jak respondent porozuměl otázkám,
- může dojít k narušení reprezentativnosti výběru tím, že skladba vrácených dotazníků nekoresponduje se skladbou základního souboru, nebo mohl za adresáta odpovědět někdo jiný.

Některé nevýhody písemného dotazování mohou být částečně nebo zcela překonány, a to především určitou znalostí charakteristik zkoumaného souboru a pečlivou přípravou aplikace techniky.

Při realizovaném sběru dat byl tazatel neustále studentům k dispozici, čímž byla vyloučena nevýhoda nekontrolovatelnosti, jak respondent porozuměl otázkám. Studenti, kteří neporozuměli otázce, měli možnost obrátit se s otázkami okamžitě na tazatele. Také fakt, že skladba vrácených dotazníků nekoresponduje se skladbou základního souboru, je v tomto případě zanedbatelný. Dotazníky byly rozdávány účelně studentům jednotlivých ročníků FES.

Při písemném dotazování je velmi důležité zabezpečit vysokou míru návratnosti dotazníků, která je těžko předvídatelná. Míra návratnosti je ovlivněna vzájemnou kombinací řady faktorů (převodní dopis, forma či obsah dotazníku, atd.). Aby nebyl ohrožen výběrový

záměr, návratnost dotazníků by měla teoreticky být 90 %. Této hranice lze za příznivých okolností dosáhnout. Skutečná v praxi se vyskytující nižší návratnost je důsledkem špatně připraveného dotazování nebo ne příliš zajímavého předmětu výzkumu [12].

Riziko špatné návratnosti dotazníků bylo eliminováno přímo sběrem dat. Studentům byly dotazníky rozdány a následně po vyplnění, přibližně po patnácti minutách opět sesbírány zpět. Tímto krokem došlo k omezení nevrácení dotazníku respondentem zpět. Problém je možné hledat pouze v chybějících hodnotách. Konkrétní hodnoty o návratnosti dotazníků jsou rozebrány v kapitole 4.1.

3.3 Klasifikace dat

Existují dvě třídy dat – kvalitativní a kvantitativní [8]:

- kvalitativní data – odlišují proměnné pomocí popisných pojmů, lze je použít pro segmentaci a klasifikaci; např. pohlaví se všeobecně klasifikuje jako „M“ coby muž a „Z“ coby žena,
- kvantitativní data – jsou charakteristická číselnými hodnotami; používají se k vytváření prediktivních modelů; např. pohlaví by mohlo být rovněž kvantitativní povahy, kdyby byla stanovena hodnotami 1 a 2 tak, že 1 = „M“ čili muž a 2 = „Z“, tedy žena.

U kvantitativních dat rozlišujeme čtyři typy [8]:

- nominální data – reprezentují kategorie neboli atributy; důležitou vlastností nominálních dat je to, že nemají relativní význam; pro modelování by měla být nominální proměnná pouze s dvěma hodnotami kódována hodnotami 0 a 1; např. muž = 1 a žena = 2 není relativní hodnota toho, že je někdo žena, dvakrát větší ani vůbec větší, než že je někdo muž,
- ordinální data - číselná data, která představují kategorie, jež mají relativní význam; lze je použít k hodnocení síly nebo důležitosti; např. student vyjádří svoji spokojenost s určitými atributy pomocí stupnice od 1 do 5; hodnota 1 vyjadřuje velkou spokojenost; hodnota 5 vyjadřuje totální nespokojenost s daným faktorem; hodnoty 2, 3 a 4 budou vyjadřovat různé úrovně spokojenosti mezi velkou spokojeností a totální nespokojeností studenta; rozdíl $5-1=4$ mezi jejich hodnoceními nemá žádný konkrétní význam,

- intervalová data – číselná data, která mají relativní význam a nemají nulový bod; zde je sčítání a odčítání bráno jako smysluplná operace; např. často jsou používány mnohem jemnější hodnoty než hodnoty od 1 do 5; typický rozsah bývá od 300 do 800; pak je možné porovnávat hodnocení měření rozdílů,
- spojitá data – nejčastějším typem dat používaných při vytváření prediktivních modelů; lze s nimi provádět všechny základní aritmetické operace včetně sčítání, odčítání, násobení a dělení; např. většina obchodních údajů, jako jsou tržby, zůstatky či minuty, jsou spojitá data.

Důležité je vyžádat si k datům potřebnou dokumentaci, například rozvržení souboru a datový slovník. Rozvržení souboru prozradí názvy proměnných, počáteční pozici dat a délku pole a typ všech proměnných. Datový slovník poskytuje informaci o typu a podrobný popis významu každé proměnné. Doporučuje se rovněž získat „výpis dat“ nebo výtisk prvních 25 – 100 záznamů. Je to nenahraditelná pomůcka pro přesvědčení, co získáváme, v případě přejetých dat [8].

3.3.1 Škály měření a typy proměnných

Odpovědím respondentů jsou přiřazeny slovní či číselné kódy. Hodnoty znaku jsou tedy kódované odpovědi neboli kategorie. Příkladem statistických znaků charakterizovaných pomocí kategorií mohou být:

- národnost (česká, slovenská,...),
- úroveň vzdělání (základní, středoškolské, vysokoškolské),
- počet dětí (0,1,2,3,...).

Uvedené příklady ilustrují různou úroveň vztahů mezi kategoriemi – v prvním případě kategorie nelze uspořádat, ve druhém je můžeme uspořádat, ve třetím můžeme vypočítat i rozdíl. V tomto smyslu můžeme hovořit o škálách měření, z nichž nejběžnější jsou nominální, ordinální, intervalová a poměrová [9].

Zvláštním typem je proměnná dichotomická (alternativní), která nabývá pouze dvou hodnot. Příkladem takových dvojic hodnot mohou být spokojen – nespokojen, ano - ne. U dichotomických proměnných můžeme dále rozlišit proměnné:

- symetrické (obě kategorie mají stejné důležitosti – muž, žena)
- asymetrické (jedna kategorie je důležitější)

U dichotomických proměnných se při výpočtech předpokládá, že jde o proměnné binární, které nabývají hodnot 0 a 1 (například číslo 1 znamená „ANO“ a číslo 0 „NE“) [9].

Každá odpověď respondenta musí být zaznamenána do proměnné (podrobněji v kapitole 4.2).

3.4 Návrh faktorů charakterizujících studenty

Jednotlivé oblasti dotazníkového šetření obsahují konkrétní otázky s daným zaměřením. Respondent odpovídá jednou z nabídnutých odpovědí s výjimkou otázek označených hvězdičkou. Na tyto označené otázky může respondent odpovědět více než jednou odpovědí.

3.4.1 Osobní dotazování

Osobní dotazování, část A v dotazníku (Příloha č. 2), je zaměřeno na základní údaje o studentovi. Základem jsou otázky zaměřené na pohlaví, věk, rodinný stav, ubytování, stravování, financování studia, volný čas a další údaje, které můžeme zařadit do oblasti osobních záležitostí studenta. Tato část obsahuje 15 otázek. Využita jsou data jak kvalitativní (M = muž, Z = žena), tak data kvantitativní (nominální a ordinální).

3.4.2 Studijní program

Část B v dotazníku specifikovaná jako Studijní program (Příloha č. 3), podává obraz o záležitostech spojených se studijním programem, střední školou, přijímacím řízením, problémovými a preferovanými oblastmi studia, využití jazykového centra, či studijních stáží a mnoho dalších otázek. Tato část obsahuje dalších 30 jednoznačných otázek. Využita jsou opět data kvalitativní (LS = letní semestr, ZS = zimní semestr) i kvantitativní (nominální – ročník studia, studijní program, studijní obor, způsob přijetí; ordinální - náročnost studia na FES, komunikační schopnosti a intervalová – platové představy studenta).

3.4.3 Hodnocení spokojenosti

Hodnocení spokojenosti, část C (Příloha č. 4), poukazuje na spokojenost studenta s aktivitami, vybavením, pedagogickým sborem či studijním oddělením FES. Data pro sestavení této části dotazníku jsou čerpána z odborné literatury [3].

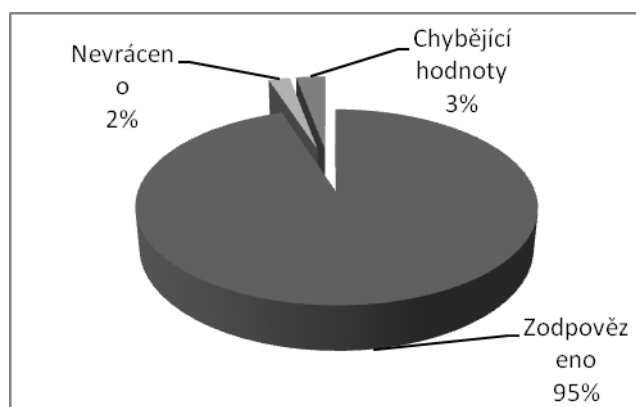
Část C obsahuje 7 okruhů otázek, na které respondent odpovídá pomocí stupnice od 1 do 5. Hodnocení 1 = velmi spokojen, 2 = spíše spokojen, 3 = spokojen, 4 = spíše nespokojen a 5 = nespokojen. Vždy je označena pouze jedna možnost z dané stupnice. Data jsou přepsána do datové matice v ordinální podobě, tak jak byla od respondentů získána.

4. Vyhodnocení dotazníkového šetření

Tato kapitola je zaměřena na zhodnocení úspěšnosti sběru dat, popis a předběžnou klasifikaci odpovědí respondentů. Pro základní analýzy je použit MS Excel.

4.1 Úspěšnost sběru dat

Jak již bylo uvedeno, sběr dat probíhal na přednáškách a seminářích jednotlivých ročníků bakalářského a navazujícího magisterského studia. Díky této metodě sběru dat je úspěšnost návratnosti téměř stoprocentní. Konkrétně z rozdaných 506 dotazníků bylo vráceno 495 zpět, což odpovídá 98% návratnosti. Z vrácených dotazníků však nebyly všechny kompletně vyplněny. V některých případech se vyskytly chybějící hodnoty, které musely být nahrazeny. Počet těchto ne zcela vyplněných dotazníků je 15, tedy opět minimální. Celková návratnost dotazníků i bez chybějících hodnot je 95 %, což je neustále velmi dobrý výsledek sběru dat (Obrázek 5).



Obrázek 5 – Úspěšnost sběru dat [vlastní]

4.2 Datová matice

Data z dotazníkového šetření nejsou zaznamenávána přímo do počítače, proto je nutné pro jejich vyhodnocení získané odpovědi převést do elektronické podoby. Odpovědi respondentů jsou tedy pomocí kódů vkládány do tabulky programu MS Excel.

Zdrojová matice je matice výchozích dat, obsahuje znaky respektive proměnné (zjištěná hodnota statistického znaku) v m sloupcích (např. typ studia, studijní program, ročník studia, pohlaví, věk, platba studia atd.) a objekty respektive vzorky v n řádcích (např. studenti jednotlivých ročníků FES), na nichž jsou tyto znaky (vlastnosti) měřeny. Řádky jsou obecně ve statistických programových systémech nejčastěji označovány jako *případy* [5]. Řádky datové matice jsou vymezeny pro odpovědi jednotlivých respondentů, sloupce obsahují odpovědi na jednotlivé otázky.

V práci jsou zpracovány tři datové matice, každá se zabývá jednou z dílčích částí dotazníkového šetření (viz kapitola 2): `Osobni_dotazovani`, `Studijni_program`, `Hodnoceni_spokojenosti`.

4.2.1 Rozvržení datového souboru `Osobni_dotazovani`

Datová matice `Osobni_dotazovani` O obsahuje 495 případů a 30 proměnných. Rozložení datové matice:

$$O = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{130} \\ x_{21} & x_{22} & \dots & x_{230} \\ \vdots & \vdots & \ddots & \vdots \\ x_{4951} & x_{4952} & \dots & x_{49530} \end{bmatrix} \quad (1)$$

Pro snadnější orientaci v datech je v Příloze č. 5 uveden datový slovník s popisem proměnných a rozsahem jejich hodnot vztahujícím se k odpovědím z dotazníku. Malou ukázkou, jak zmiňovaný datový slovník vypadá je následující Tabulka č. 1.

Tabulka 1 – Ukázka datového slovníku Osobni_dotazovani [vlastní]

Název atributu	Datový typ	Rozsah hodnot	Kód - Hodnota atributu		Stručný popis
x1	Flag	Z/M	M	Muž	pohlaví
			Z	Žena	
x2	Range	[19 - 24]	19	19	věk
			20	20	
			21	21	
			22	22	
			23	23	
			24	24 a více	
x3	Set	[1,2,3]	1	Svobodný/á	rodinný stav
			2	Ženatý/vdaná	
			3	Rozvedený/á	
x4	Set	[1 - 15]	1	Karlovarský kraj	kraj
			2	Plzeňský kraj	
			3	Ústecký kraj	
			4	Jihočeský kraj	
			5	Liberecký kraj	
			6	Středočeský kraj	
			7	Praha	
			8	Královehradecký kraj	
			9	Pardubický kraj	
			10	Vysočina	
			11	Jihomoravský kraj	
			12	Olomoucký kraj	
			13	Moravskoslezský kraj	
			14	Zlínský kraj	
			15	Slovensko	

4.2.2 Rozvržení datového souboru Studijni_program

Datová matice S (Studijni_program) obsahuje 51 proměnných, přičemž počet případů zůstává neustále 495 jako u předchozí datové matice. Číslování proměnných zde navazuje na číslování proměnných matice O. Začíná tedy proměnnou x_{31} a poslední proměnná je označena x_{82} . Důvodem navázání číslování je záměr propojení těchto matic při zpracování následných analýz. Její rozvržení vypadá následovně:

$$S = \begin{bmatrix} x_{131} & x_{132} & \dots & x_{182} \\ x_{231} & x_{232} & \dots & x_{282} \\ \vdots & \vdots & \ddots & \vdots \\ x_{49531} & x_{49532} & \dots & x_{49582} \end{bmatrix} \quad (2)$$

Datový slovník s názvy proměnných, hodnotami a stručným popisem je součástí přílohy práce (Příloha č. 6). Ukázka datového slovníku je zobrazena v Tabulce č. 2.

Tabulka 2 - Ukázka datového slovníku Studijni_program [vlastní]

Název atributu	Datový typ	Rozsah hodnot	Kód - Hodnota atributu		Stručný popis
x31	Flag	Prezencni/Kombinovane	Prezencni	Prezenční studium	forma studia
			Kombinovane	Kombinované studium	
x32	Flag	1/2	1	Bakalářský	typ studia
			2	Navazující	
X33	Ordered Set	[1,2,3]	1	První	ročník studia
			2	Druhý	
			3	Třetí	
			4	Čtvrtý	
			5	Pátý	
x34	Set	[EM, HPS,SII]	HPS	Hospodářská politika a správa	studijní program
			EM	Ekonomika a management	
			SII	Systémové inženýrství a informatika	
x35	Set	[EMP,..., VES]	VES	Veřejná ekonomika a správa	studijní obor
			EK	Ekonomika pro kriminalisty	
			ECS	Ekonomika celní správa	
			EVS	Ekonomika veřejného sektoru	
			RR	Regionální rozvoj	
			RVE	Regionální a veřejná ekonomie	
			MP	Management podniku	
			MOPS	Management ochrany podniku a společnosti	
			EPP	Ekonomika a provoz podniku	
			EMP	Ekonomika a management podniku	
			IVS	Informatika ve veřejné správě	
			RIM	Regionální a informační management	
			IBS	Informační a bezpečnostní systémy	
PI	Pojistné inženýrství				

4.2.3 Rozvržení datového souboru Hodnoceni_spokojenosti

Datová matice H (Hodnoceni_spokojenosti) obsahuje opět 495 případů, na které je propojeno 50 proměnných. Proměnné jsou označeny písmenem S. Rozvržení matice vypadá následovně:

$$H = \begin{bmatrix} S_{11} & S_{12} & \dots & S_{1\ 50} \\ S_{21} & S_{22} & \dots & S_{2\ 50} \\ \vdots & \vdots & \ddots & \vdots \\ S_{4951} & S_{4952} & \dots & S_{495\ 50} \end{bmatrix} \quad (3)$$

Odpovědi respondentů v této matici jsou zaznamenávány přímo ve formě škál, zodpovězených studenty.

Získaná data z této matice nebudou blíže analyzována. V Příloze č. 7 je uvedena pouze ukázka matice H.

4.3 Analýza proměnných

Mezi základní typy analýz patří zjištění rozdělení četností různých variant hodnot pro každý sledovaný znak a výpočet souhrnných charakteristik. Nejprve je sledován celý základní soubor, do šetření jsou tedy zahrnuti všichni dotázaní respondenti. Poté je analýza zaměřena na konkrétní části studentů, je provedena analýza na výběrovém souboru. Na základě získaných výsledků je proveden úsudek o celém souboru [9].

4.3.1 Absolutní četnost

Počet statistických jednotek, patřících do určité třídy, nazýváme její absolutní četností. Při každém třídění je nutné dodržovat zásady úplnosti a jednoznačnosti. Zásada úplnosti vyžaduje vytvoření takových tříd, aby každá statistická jednotka mohla být zařazena do některé z nich. Zásada jednoznačnosti požaduje, aby každá statistická jednotka byla zařazena jen v jedné třídě [4].

Pokud zkoumáme ve statistickém souboru kvantitativní znak, pak provádíme třídění podle velikosti hodnot tohoto znaku. V případě kvalitativního znaku představuje každá jeho hodnota samostatnou třídu, nelze je uspořádat podle velikosti ani rozdělit do intervalových tříd. Absolutní četnost m_i znamená, kolikrát byla hodnota x_i naměřena. Platí:

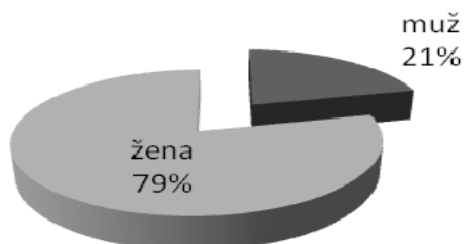
$$\sum_{i=1}^r m_i = n \quad (4)$$

kde n je rozsah náhodného výběru a r je počet tříd. Rozdělení četností může být reprezentováno ve dvou formách - v tabulce nebo grafu [4].

4.3.2 Prvotní analýza

Vzhledem k velkému rozsahu dat je potřebné pro první představu realizovat základní analýzu dat. Ze základního souboru lze vyčíst hned několik základních faktů. Jedná se především o statistické zastoupení studentů jednotlivých ročníků, studijních programů a zastoupení mužů a žen, jak je uvedeno v následujícím přehledu:

- počet mužů a žen,



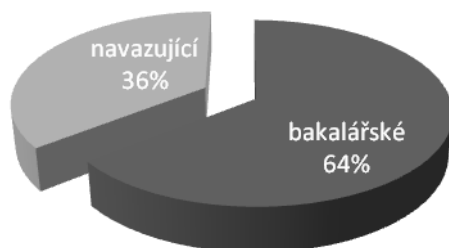
Obrázek 6 – Zastoupení pohlaví [vlastní]

- zastoupení studentů z jednotlivých krajů,

Tabulka 3 – Zastoupení studentů v krajích [vlastní]

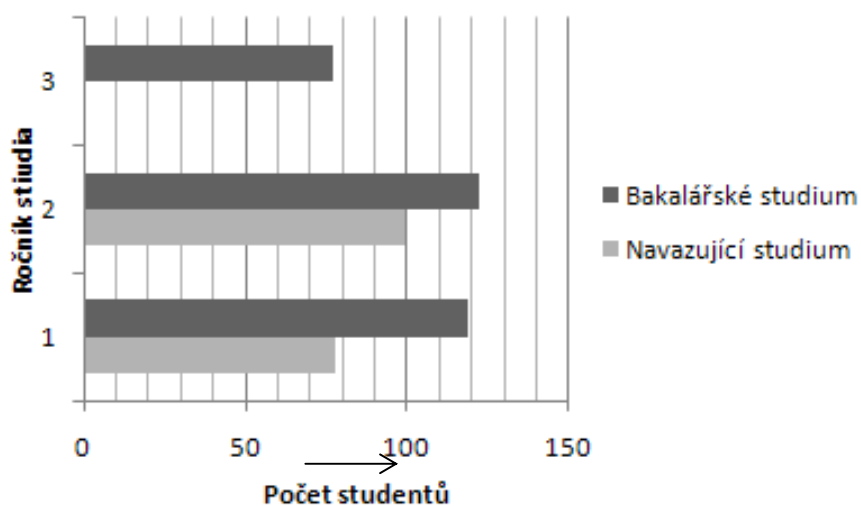
Název kraje	Četnosti	Procenta
Karlovarský kraj	4	0,81
Plzeňský kraj	11	2,22
Ústecký kraj	4	0,81
Jihočeský kraj	2	0,40
Liberecký kraj	15	3,03
Středočeský kraj	63	12,73
Praha	5	1,01
Královéhradecký kraj	108	21,82
Pardubický kraj	200	40,40
Vysočina	37	7,47
Jihomoravský kraj	13	2,63
Olomoucký kraj	16	3,23
Moravskoslezský kraj	4	0,81
Zlínský kraj	4	0,81
Slovensko	1	0,20

- zastoupení studentů bakalářských a navazujících studijních programů,



Obrázek 7 – Počty studentů prezenčního a kombinovaného studia [vlastní]

- zastoupení jednotlivých ročníků,



Obrázek 8 – Počty studentů jednotlivých ročníků [vlastní]

- zastoupení studijních programů a oborů,

Tabulka 4 – Počet studentů jednotlivých oborů [vlastní]

Program	Obor	Absolutní četnost	Relativní četnost [%]
HPS	VES	216	45
	EVS	47	9
	EK	0	0
	ECS	0	0
	RR	0	0
EM	MP	56	11
	MOPS	14	3
	EPP	29	6
	EMP	46	9
SII	IVS	36	7
	RIM	30	6
	IBS	7	1
	PI	14	3

- navštěvované kluby v Pardubicích.

Tabulka 5 – Oblíbené kluby v Pardubicích [vlastní]

Kluby v Pardubicích	Počet studentů
Žádný klub	204
Dýdý baba	137
Áčko	45
Žlutý pes	23
Hany Bany	19
Bum bum kafe	7
Klec	6
Pat a Puf	4

Po zhlédnutí výsledků lze jednoznačně charakterizovat základní soubor, který je tvořen studenty FES. Značnou převahu mají mezi studenty ženy (Obrázek 6). Čtyřicetiprocentní převahu mají studenti z Pardubického kraje, následuje Královéhradecký kraj s 22% zastoupením. Na druhou stranu nejméně zastoupená kategorie je Slovensko s jedním studentem (Tabulka 3).

Předpokládaný stav ukazuje Obrázek 7 zobrazující počty studentů typů studia. Početnější skupinou je zde s 64 procenty zastoupen bakalářský typ studia. Obdobný je i výsledek Obrázku 8, dokumentující počty studentů jednotlivých ročníků. Největší zastoupení mají studenti prvního a druhého ročníku bakalářského studia. Pro následující analýzu mají však

větší vypovídací schopnost respondenti vyšších ročníků. Analýza ukazuje výskyt téměř 80 studentů prvního a 100 studentů druhého ročníku navazujícího studia. V závěru dotazníku byla položena otázka zda-li studenti navštívili psychologickou poradnu ve škole v případě potřeby. Z dotázaných studentů kladně odpovědělo pouze 11.

Následující Tabulka 4 ukazuje počet studentů jednotlivých studijních oborů. Alarmujících je zde hned několik prvků. Jedním, maximálním prvkem jsou studenti programu Hospodářská politika a správa (HPS) oboru Veřejná ekonomika a správa (VES) se zastoupením 216 studentů. Druhým alarmujícím prvkem je naopak minimální počet studentů programu Systémoví inženýrství a informatika (SII) oboru Informační a bezpečnostní systémy (IBS). Dalšími nepříliš pozitivními fakty jsou nulové zastoupení oborů Ekonomika pro kriminalisty (EK), Ekonomika a celní správa (ECS) a Regionální rozvoj (RR). Vzhledem k tomu, že výběr studentů byl uskutečněn záměrně na přednášky a semináře všech oborů, vypovídají tato čísla o žádaných a naopak méně žádaných oborech FES.

Spíše jako zajímavost byly v dotazníku zjišťovány kluby, které studenti v Pardubicích navštěvují. Celkem zajímavé je, že studenti ve 204 případech odpověděli, že v Pardubicích žádný klub nenavštěvují. Mezi studenty, kteří navštěvují pardubické kluby je nejvíce oblíben studentský klub Dýdy baba či Áčko. Ostatní kluby navštěvované pardubickými studenty jsou uvedeny v Tabulce 5.

Spokojenost respondentů s univerzitními koleji

Odpověď na otázku, kde studenti bydlí v Pardubicích má 3 varianty – Vysokoškolské koleje, Privát či byt v Pardubicích nebo možnost že studenti v Pardubicích nebydlí a denně dojíždí. Z dotázaných 495 studentů 197 denně dojíždí a 190 bydlí v Pardubicích mimo vysokoškolské koleje. Menšina studentů – 108 bydlí na kolejích (Tabulka 6). Tento fakt může vypovídat v prvním případě o tom, že studenti jsou z blízkého okolí, nebo že úroveň kolejí je nevyhovující.

Tabulka 6 – Ubytování studentů [vlastní]

Ubytování v Pardubicích	Počet studentů	
	Absolutní četnost	Relativní četnost [%]
Vysokoškolské koleje	108	22
Privát	190	38
Dojíždí	197	40

O absolutní spokojenosti studentů s kolejemi, tedy s cenou, službami, systémem přidělení kolejí a způsobem propojení účtu kolejí a menzy vypovídá následující analýza. Spokojeni jsou pouze dva studenti ze 108, kteří na kolejích bydlí. Tento celkem alarmující fakt ukazuje Obrázek 9.

Value	Proportion	%	Count
Nespokojen		98,15	106
Spokojen		1,85	2

Obrázek 9 – Spokojenost s vysokoškolskými kolejemi [vlastní]

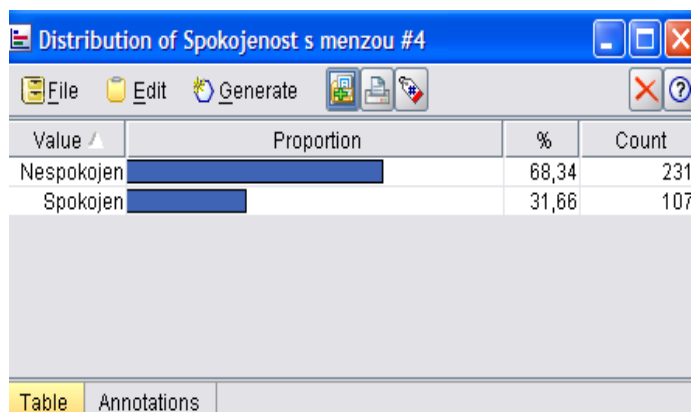
Spokojenost studentů podle jednotlivých dílčích otázek je již uspokojivější (Tabulka 6).

Tabulka 7 – Spokojenost s univerzitními kolejemi [vlastní]

Cílový atribut	Spokojen	Nespokojen	Neodpověděl
Cena kolejí	18	86	4
Cena = služby kolejí	13	91	4
Systém přidělení kolejí	63	42	3
Propojení účtů kolejí a menzy	64	41	3

Spokojenost respondentů se službami menzy

Obdobná analýza byla provedena i na spokojenost studentů s menzou. Z oslovených 495 studentů využívá služeb stravování v menze 338 studentů. Hodnoceny byly otázky ohledně ceny a kvality jídla, otvírací doby menzy, systému volného výběru jídel a to využívají-li studenti nabídky doplňkového sortimentu. Výsledek absolutní spokojenosti s cenou, kvalitou jídla, otvírací dobou a systémem výběru jídel je již uspokojivější než u předchozího zkoumání. Spokojeno je 107 (32 %) respondentů, nespokojeno je zbylých 231 (68 %).



Obrázek 10 – Spokojenost s menzou [vlastní]

Spokojenost podle jednotlivých dílčích otázek ukazuje Tabulka č. 8. Ze všech studentů, kteří využívají služeb menzy, jeden student odmítl odpovědět na dílčí otázky. Ostatní hodnotili následovně:

S cenou jídla je 201 studentů spokojeno, 136 je nespokojeno. S faktem, že cena jídla odpovídá jeho kvalitě, souhlasí 150 studentů. Většina studentů však tento fakt popřela, tedy 187 studentů je nespokojeno s kvalitou jídla na této cenové úrovni. Naprostá většina (283) oslovených studentů souhlasí s otvírací dobou, pouhých 54 se vyslovilo proti. Také skutečnost, že si jídla studenti vybírají bez předchozího objednání je pro většinu (309) studentů uspokojivá. Pouhým 28 studentům systém volného výběru jídel v menze nevyhovuje.

Tabulka 8 – Spokojenost studentů s menzou [vlastní]

Cílový atribut	Spokojen	Nespokojen	Neodpověděl
Cena jídla	201	136	1
Cena = kvalita jídla	150	187	1
Otvírací doba menzy	283	54	1
Systém výběru jídel	309	28	1

Opakování předmětů

V úvodní části byly již analyzovány základní znaky popisující studenty (typ studia, studijní program, ročník studia, atd.). Kromě těchto obecných otázek byly v dotazníku (části B) dotazováni studenti specifickými otázkami. Jedna z nich je Opakování předmětů.

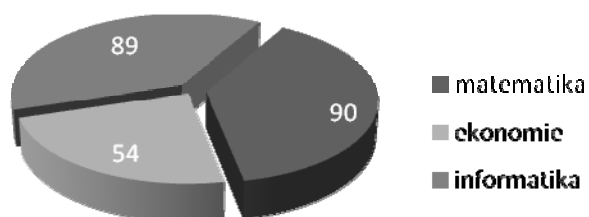
Tabulka 9 - Počet studentů opakujících předmět [vlastní]

Opakování	Počet studentů
ANO	162
NE	333

Tabulka 10 – Počet opakovaných předmětů [vlastní]

Počet opakovaných předmětů	Počet studentů
Jeden	81
Dva	58
Tři	23

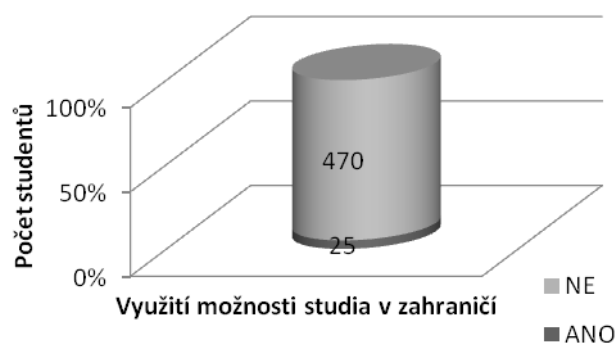
Celkový počet studentů, kteří opakují předměty z loňských let je 162 z oslovených studentů (Tabulka 9). Tabulka 10 pak ukazuje, kolik studentů opakuje jeden, dva nebo tři předměty. Specifikovány byly otázky na oblasti předmětů, které studenti opakují. Jak je vidět na Obrázku 11, nejvíce studentů opakuje předměty z oblasti matematika a informatika. Nejméně opakovaných předmětů je v oblasti ekonomie.



Obrázek 11 – Počet studentů opakujících jednotlivé oblasti [vlastní]

Využití možnosti studia v zahraničí

Další specifická otázka je zaměřena na studium v zahraničí. Z celkového počtu 495-ti oslovených studentů využilo tuto možnost pouhých 25 studentů (Obrázek 12). Tento minimální počet studentů, kteří vycestovali do zahraničí, by mohl být impulsem pro změny na FES, motivující studenty zúčastnit se zahraničních pobytů na partnerských univerzitách.



Obrázek 12 – Studium v zahraničí [vlastní]

Skupinu navštívených destinací seřazených od nejžádanějších po méně navštěvované zobrazuje Tabulka 11.

Tabulka 11 – Destinace navštěvované studenty [vlastní]

Země	Počet studentů
Portugalsko	14
Finsko	3
Německo	2
Itálie	1
Madeira	1
Slovensko	1
Slovinsko	1
Turecko	1

Všichni studenti vycestovali v rámci programu Erasmus kromě jednoho studenta, který využil jiný nejmenovaný program. Studium absolvovali studenti především v anglickém jazyce. V Německu studovali studenti v německém jazyce. U některých studentů byly zmíněny i národní jazyky zemí kde studovali, ale na prvním místě byl vždy anglický jazyk. Pobyt neměl přínos pouze u jednoho studenta, který absolvoval pobyt na Madeiře. Pro všechny ostatní byl tento pobyt přínosný. Většina studentů vycestovala na letní semestr - 16 studentů, na zimní semestr vycestovalo 8 studentů. Pouze jeden student zůstal v zahraničí na oba semestry.

Využívání služeb Univerzitní knihovny

Jedna z otázek je směřována na využívání Univerzitní knihovny (UK). Obecně by se dalo předpokládat, že každý student FES bude využívat univerzitní knihovnu. Jak je vidět z Tabulky 11, objevilo se 25 studentů, kteří ke svému studiu UK nepotřebují. Otázka byla blíže specifikována, za jakým účelem studenti knihovnu navštěvují, jak často a jak jsou spokojeni s otvírací dobou a poskytovanou literaturou.

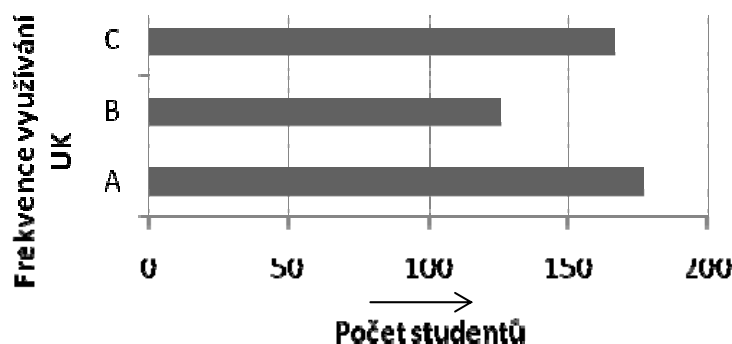
Tabulka 12 - Využívání Univerzitní knihovny [vlastní]

Využívání UK	Absolutní četnost	Relativní četnost [%]
ANO	470	95
NE	25	5

Tabulka 13 – Účel návštěvy Univerzitní knihovny [vlastní]

Účel návštěvy	Absolutní četnost	Relativní četnost [%]
Studium	423	85
Internet, PC	156	32
Denní tisk, časopisy	70	14

Jak je patrné z Tabulky 12 nejvíce (85 % studentů) je UK využívána za účelem studia. Za účelem použití internetu a počítačů je knihovna využívána pouze 32% studentů. Četba denního tisku a časopisů je na posledním místě se čtrnáctiprocentním zastoupením.



Obrázek 13 – Frekvence využívání Univerzitní knihovny [vlastní]

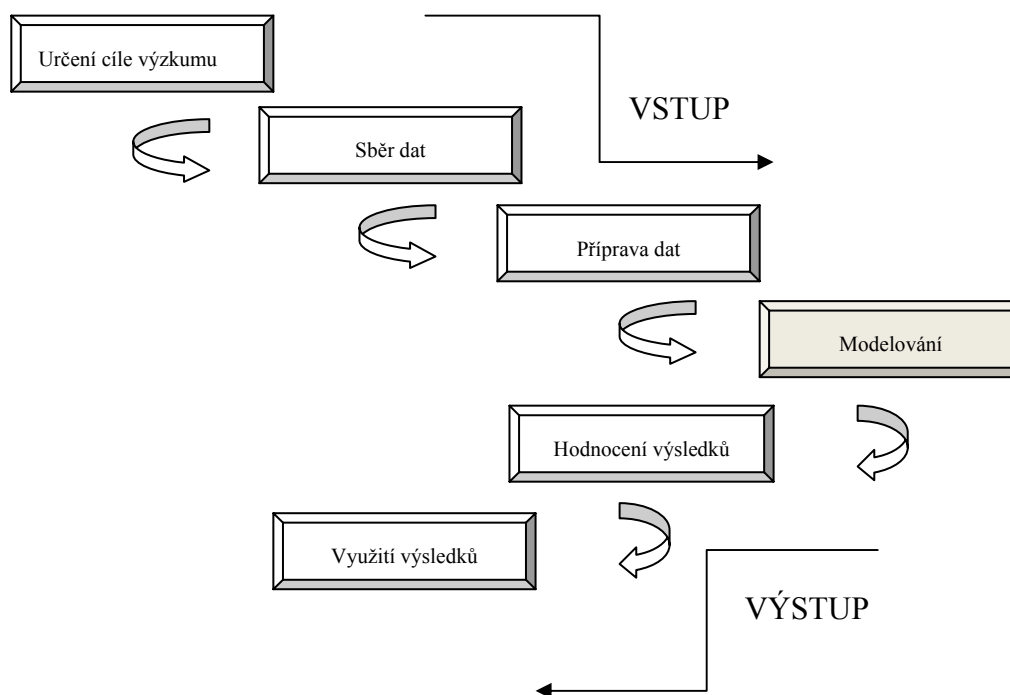
Jedenkrát týdně (A) a jedenkrát měsíčně (C) studenti využívají UK nejvíce. Vícekrát v týdnu (B) je již méně časté u studentů chodit do knihovny.

5. Návrh klasifikačního modelu a popis metod

Tato kapitola je zaměřena na obecný model procesu klasifikace a následný rozbor metod, v nichž je vyjadřována závislost určité (vysvětlované) proměnné na několika jiných (vysvětlujících) proměnných [9].

5.1 Obecné dobývání znalostí

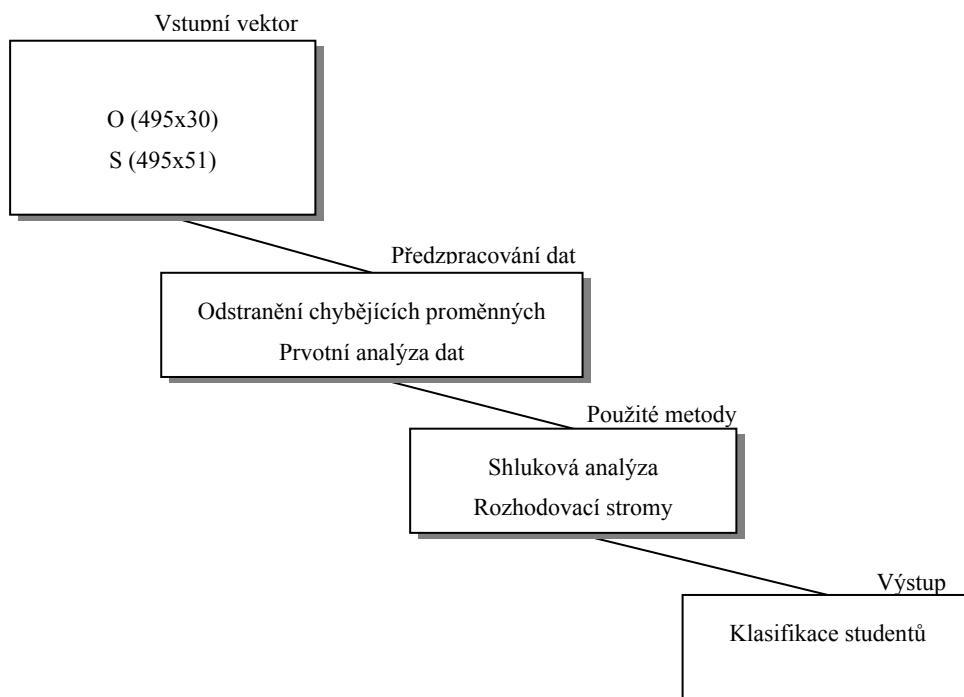
Jádrem celého procesu dobývání znalostí je použití analytických metod. Základními prvky jsou pro analytické metody (modelování) vstup a výstup. Vstupy jsou v případě této práce atributy datových matic O a S, tedy data získaná dotazníkovým šetřením. Datová matice O je u některých modelů analyzována společně s maticí S. Propojením a následným zpracováním vzniknou výstupy, klasifikující studenty FES, tedy určité znalosti [8].



Obrázek 14 – Začlenění modelování do dobývání znalostí, upraveno podle[1]

5.2 Schéma modelu

Samotný model je hlavní součástí tvorby výstupu. Jedná se o vývojový algoritmus zahrnující následující hlavní části: vstupní vektor, předzpracování, použité metody a výstup. Jak je vidět na Obrázku 15, jednotlivé fáze na sebe volně navazují.



Obrázek 15 – Popis modelu [vlastní]

5.3 Metody klasifikace

Nejjednodušší je modelování v případě dichotomické vysvětlované proměnné. Při znalosti hodnot vysvětlujících proměnných se pak objekt (respondent) zařazuje do určité skupiny - klasifikuje. Počet skupin je dán počtem kategorií vysvětlované proměnné. Pro tyto situace byly navrženy různé přístupy [9].

Dolování dat z databází, je založeno na množství matematických a statistických technik: rozhodovací stromy, neuronové sítě, genetické algoritmy, shlukování a klasifikace [1]. V této práci jsou použity zejména rozhodovací stromy a shlukování. Vstupem pro dané modely jsou data nejen v podobě dichotomických proměnných, ale i nominální a ordinální.

5.3.1 Rozhodovací stromy

Rozhodovací strom (RS) je prediktivní model, který zobrazuje data v podobě stromu. Každý uzel určuje kritérium pro následné rozdělení dat do jednotlivých větví. Strom tak rozděluje veškerá zdrojová data do segmentů, kde každý list odpovídá určitému segmentu definovanému předchozími uzly. Data, která jsou zařazena do určitého segmentu, se vyznačují shodnými vlastnostmi. Rozhodovací stromy mohou být založeny na množství algoritmů. Příklady algoritmů: CHAID (Chi-squared Automatic Interaction Detector), QUEST (The Quick, Unbiased, Efficient Statistical Tree), C&RT (Classification and Regression Trees), C5.0 [8].

Schéma se začíná tvořit od kořenového uzlu, který představuje všechny sledované respondenty. V tomto uzlu se vyznačí četnost variant odpovědí u vysvětlované proměnné. Zkoumáme postupně vysvětlující proměnné a pomocí věrohodnostního poměru zjišťujeme, na které je vysvětlující proměnná nejvíce závislá. V případě, kdy je vysvětlující proměnná ordinální, se prověřují též nové proměnné s menším počtem kategorií, které vznikly sloučením některých původních kategorií [9].

Algoritmus pro tvorbu RS [1]:

- zvol jeden atribut jako kořen dílčího stromu,
- rozděl data v tomto uzlu na podmnožiny podle hodnot zvoleného atributu a přidej uzel pro každou podmnožinu,
- existuje-li uzel, pro který nepatří všechna data do téže třídy, pro tento uzel opakuj postup od bodu 1, jinak skonči.

Při výběru vhodného algoritmu je cíl zřejmý: vybrat takový atribut, který od sebe nejlépe odliší příklady různých tříd. Vodítkem pro volbu jsou charakteristiky atributu převzaté z teorie informace nebo pravděpodobnosti: entropie, informační zisk, poměrný informační zisk, χ^2 , F-test nebo Giniho index [1].

Entropie je pojem používaný v přírodních vědách pro vyjádření míry neuspořádanosti nějakého systému. V teorii informace je entropie definována jako funkce:

$$H = -\sum_{t=1}^T (p_t \log_2 p_t) \quad (5)$$

kde p_t - pravděpodobnost výskytu třídy t ; T – počet tříd.

Je-li $p = 1$ (všechny případy patří do jedné třídy) nebo $p = 0$ (žádný příklad nepatří do této třídy), je entropie nulová. Jsou-li obě třídy zastoupeny stejným počtem příkladů ($p = 0,5$), je entropie maximální.

Výpočet entropie pro jeden atribut se provádí následovně. Pro každou hodnotu v , kterou může nabýt uvažovaný atribut A spočítej podle uvedeného vzorce entropii $H(A(v))$ na skupině příkladů, které jsou pokryty kategorií $A(v)$.

$$H(A(v)) = -\sum_{t=1}^T \frac{n_t(A(v))}{n(A(v))} \log_2 \frac{n_t(A(v))}{n(A(v))} \quad (6)$$

Spočítá se střední entropie $H(A)$ jako vážený součet entropií $H(A(v))$, přičemž váhy v součtu jsou relativní četnosti kategorií $A(v)$ v datech D_{TR} .

$$H(A) = -\sum_{v \in Val(A)} \frac{n(A(v))}{n} H(A(v)). \quad (7)$$

Pro větvení stromu se poté vybere atribut s nejmenší entropií [1].

Informační zisk (informatik gain) i poměrový informační zisk (informatik gain ratio) jsou míry odvozené z entropie. Informační zisk (IZ) se spočítá jako rozdíl entropie pro celá data (pro cílový atribut) a pro uvažovaný atribut. IZ tak měří redukci entropie způsobenou volbou atributu A :

$$Zisk(A) = H(C) - H(A), \text{ kde } H(C) = -\sum_{t=1}^T \frac{n_t}{n} \log_2 \frac{n_t}{n} \quad (8)$$

V případě IZ hledáme atribut s maximální hodnotou [1].

Někdy se používá jako kritérium pro volbu atributu **poměrný informační zisk**, který kromě entropie bere do úvahy i počet hodnot atributu. Výpočet vypadá následovně [1]:

$$\text{Poměrný zisk } (A) = \frac{\text{Zisk}(A)}{\text{Větvení}(A)} \quad (9)$$

kde

$$\text{Větvení } (A) = - \sum_{v \in \text{Val}(A)} \frac{n(A(v))}{n} \log_2 \frac{n(A(v))}{n}. \quad (10)$$

Gini index může mít stejnou roli, jakou hrála v předchozích úvahách entropie. Tento index se spočítá jako:

$$\text{Gini} = 1 - \sum_{t=1}^T (p^2 t) \quad (11)$$

kde p_t je relativní počet příkladů t-té třídy.

Hodnota indexu je minimální v případě, že příklady patří do jedné ze tříd a maximální v případě, že příklady jsou rovnoměrně rozděleny mezi obě třídy.

Hodnotu Giniho indexu pro jednu hodnotu spočítáme analogicky jako hodnotu entropie pro jeden atribut. Tedy tak, že pro každý atribut spočítáme vážený součet indexu pro jednotlivé hodnoty atributu, přičemž váhy budou relativní četnosti příslušných hodnot:

$$\text{Gini}(A) = \sum_{v \in \text{Val}(A)} \frac{n(A(v))}{n} \text{Gini}(A(v)) \quad (12)$$

$$\text{Gini}(A) = 1 - \sum_{t=1}^T \left(\frac{n_t(A(v))}{n(A(v))} \right)^2 \quad (13)$$

Pro větvení použijeme atribut, který bude mít nejmenší hodnotu tohoto indexu [1], [2].

Jako kritérium pro volbu atributu je možné použít i χ^2 test. Je to míra, umožňující vyhodnocovat vzájemné souvislosti mezi dvěma atributy. Tento test je založen na vyhodnocení rozdílu mezi pozorovanými četnostmi jednotlivých kombinací a četnostmi očekávaných při platnosti hypotézy o nezávislosti obou veličin [1]. Testujeme statistickou hypotézu, že náhodný výběr je utvořen ze základního souboru s určitým konkrétním rozdělením pravděpodobností.

Testovací kritérium má tvar:

$$\chi = \sum_{i=1}^r \frac{(m_i - np_i)^2}{np_i} \quad (14)$$

Kde jednotlivé symboly značí:

r – počet tříd, do kterých byl zatříděn náhodný výběr,

m_i - absolutní četnost i -té třídy (počet hodnot náhodné veličiny v dané třídě),

n – rozsah náhodného výběru,

p_i – pravděpodobnost i -té třídy v předpokládaném rozdělení pravděpodobností,

np_i – teoretická četnost i -té třídy.

Třídy je třeba určit takovým způsobem, aby platilo $\sum_i p_i = 1$ [4].

Při volbě vhodného atributu pro větvení stromu se vybere atribut, který nejvíce souvisí s cílovým atributem - χ^2 má největší hodnotu [1].

Fisherův F-test neboli dvouvýběrový test významnosti pro rozptyl je použit u modelů CHAID a QUEST [7]. Předpokládejme, že náhodný výběr $(X_1, X_2, \dots, X_{n_1})$ je vytvořen ze základního souboru X s $N(\eta_1, \sigma_1)$ rozdělením pravděpodobností, náhodný výběr $(Y_1, Y_2, \dots, Y_{n_2})$ je vytvořen ze základního souboru Y s $N(\eta_2, \sigma_2)$ rozdělením pravděpodobností. Dále je předpokládáno, že náhodné veličiny $X_1, X_2, \dots, X_{n_1}, Y_1, Y_2, \dots, Y_{n_2}$ jsou nezávislé (tj. výběry jsou nezávislé). Cílem je ověřit platnost hypotézy o rovnosti disperzí souborů, což je symbolicky zapsáno: $H_0 = DX = DY$ proti alternativní hypotéze $H_1 = DX \neq DY$.

Testovací kritérium je náhodná veličina:

$$F = \frac{s_1^{-2}}{s_2^{-2}} \quad (15)$$

Za předpokladu platnosti hypotézy má náhodná veličina F Fisher – Snedecorovo rozdělení pravděpodobností s $n_1 - 1$ a $n_2 - 1$ stupni volnosti. Kritická oblast je definována vztahem: $F > F_1$ nebo $F > F_2$. Kritické hranice F_1 , F_2 jsou vypočítány následujícím způsobem:

$$P(F < F_1) = \frac{\alpha}{2} \qquad P(F < F_1) = \frac{\alpha}{2} \qquad (16)$$

$$F_1 = F_{F(n_1-1, n_2-1)}^{-1}\left(\frac{\alpha}{2}\right) \qquad F_2 = F_{F(n_1-1, n_2-1)}^{-1}\left(\frac{2-\alpha}{2}\right)$$

Ve většině tabulek Fisher-Snedecorova rozdělení pravděpodobností jsou tabelovány jen pravé kritické hranice, proto je výhodnější používat testovací kritérium v tvaru:

$$F = \frac{\max\{s_1^{-2}; s_2^{-2}\}}{\min\{s_1^{-2}; s_2^{-2}\}} \qquad (17)$$

V tomto případě padnou do kritické oblasti jen velké hodnoty testovacího kritéria F , proto kritická oblast bude tvořená množinou těch hodnot testovacího kritéria F , pro které platí $F > F_2$. Hladina významnosti i v tomto případě zůstává α .

Kritickou oblast zapisujeme [4]:

$$W = \{F : F > F_2\} \text{ kde } F_2 = F_{F(n_1-1, n_2-1)}^{-1}\left(\frac{2-\alpha}{2}\right). \qquad (18)$$

Cílem analýzy pomocí RS je zjistit závislosti mezi výslednou proměnnou (vysvětlovanou) a ostatními proměnnými (vysvětlující proměnné). Z datových matic O a S jsou vybrány cílové atributy, které budou zkoumány. Patří mezi ně, v návaznosti na počáteční analýzu, spokojenost s cenou jídla v menze a parkování v okolí univerzity. Dále bude zkoumáno opakování předmětů – opakování předmětů týkajících se matematiky, opakování předmětů informatiky, studium v zahraničí a nakonec budou vyhodnoceny závislosti s proměnnou vytyčené cíle studentů.

Použité modely rozhodovacích stromů: CHAID, QUEST, C&RT, C5.0.

CHAID

Algoritmus CHAID patří mezi nejrozšířenější rozhodovací stromy (RS). Tento RS začíná u celého souboru a dochází k postupnému větvení souboru. Z jednoho uzlu může vycházet libovolný počet větví (nebinární RS). Algoritmus CHAID je zařazen mezi rekurzivní – každý uzel se dělí podle stejného předpisu a zastaví se tehdy, neexistuje-li příznačné rozdělení. Vstupy mohou být číselné i kategorizované, stejně tak i výstupy. Použití statistického testu závisí na typu cílového atributu. F test se používá v případě, že cílový atribut je spojitý, χ^2 test v případě nominální proměnné a poměrný test u ordinální proměnné [1], [2], [7].

QUEST

Rozhodovací strom QUEST má celkem jednoduchý výpočet. Dává přednost binárnímu dělení. Jednoznačnou předností tohoto RS je jeho nezaujatost výběru proměnných a jeho rychlost výpočtu. Při výpočtu používá mezi vstupní a cílovou proměnnou F test nebo Levenson test (pro ordinální a spojitě proměnné) nebo χ^2 test (pro nominální proměnné) [1], [2], [7].

C&RT

Tento RS umožňuje pouze binární stromy – stromy u kterých jsou možné pouze dvě větve. Atribut pro větvení stromu je vypočten pomocí Gini indexu – atribut s nejmenší hodnotou tohoto indexu. Cílová veličina je předem definována. Na vstupu i na výstupu mohou být jak číselná data, tak kategorizovaná. Strom vytvoří podmnožiny dat tak, aby byla stejnorodá s ohledem na definovanou cílovou proměnnou. Umožňuje nastavení maximální úrovně hloubky [1], [2], [7].

C5.0

Algoritmus založený na poměrném informačním zisku - pro větvení je vybrán atribut s největším informačním ziskem. Rozhodovací strom umožňuje i nebinární stromy – jsou možné i více než dvě větve. Algoritmus umí pracovat s číselnými i kategorickými daty, na výstupu musí být ale kategorická data [1], [2], [7].

5.3.2 Shluková analýza

Shlukování je označováno jako jeden ze základních typů získávání znalostí. Základním cílem shlukové analýzy je zařadit objekty do skupin (shluků) a to především tak, aby dva objekty stejného shluku si byly více podobné, než dva objekty z různých shluků. Přitom objekty mohou být různého charakteru [10].

Analýza shluků se hodí zejména tam, kde objekty projevují přirozenou tendenci seskupovat se. Podle způsobu shlukování se postupy dělí na hierarchické a nehierarchické. Hierarchické se dělí dále na aglomerativní a divizní [5].

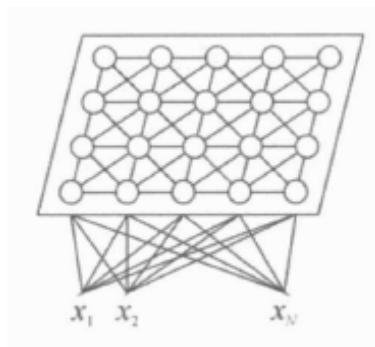
K posuzování vztahů mezi kategoriemi slouží míry podobnosti, resp. nepodobnosti. Při hierarchickém shlukování se obvykle vytváří matice vzdáleností (nepodobností), v níž je každá dvojice kategorií charakterizována číslem vyjadřujícím, do jaké míry jsou sledované kategorie blízké. Čím více se míra podobnosti blíží hodnotě 0, tím jsou si kategorie bližší. A naopak, čím míra nepodobnosti dosahuje vyšší hodnoty, tím jsou kategorie rozdílnější [9].

Cílem shlukové analýzy v této práci je rozdělit studenty do skupin a zjistit jakými studenty jsou skupiny tvořeny. Vše je realizováno na propojených datových maticích O a S. Do shlukové analýzy jsou zařazeny bez rozdílu všechny atributy. Oproti RS zde není určena cílová proměnná, jde tedy o komplexní analýzu celého datového souboru, která bude provedena na závěr práce. Použita je jedna z metod shlukové analýzy - Kohonenovy mapy.

Kohonenovy mapy

Kohonenovy mapy jsou jedním z modelů neuronových sítí vhodných pro analýzu dat. Někdy jsou nazývány jako samoorganizující se mapy (SOM). První model SOM byl navržen v roce 1982, jde o model učení bez učitele. Jedná se o nehierarchické shlukování s proměnným počtem shluků. Po rozpoznání shluků v datech lze tuto síť využít pro klasifikaci [10].

Síť SOM má dvě vrstvy, a to vstupní a výstupní vrstvu neuronů (známou také pod názvem Kohonenova vrstva, nebo vrstva topologické mapy). Jednotky ve vrstvě topologické mapy jsou prostorově uspořádány - typicky ve dvou dimenzích. Schéma neuronové sítě je znázorněno na Obrázku 16. Spodní vstupní vrstva je tvořena N neurony. Do této vrstvy přicházejí vstupní vektory, které budou transformovány. Síť se učí rozpoznávat podobné vstupy podle algoritmu [10].



Obrázek 16 – Schéma neuronové sítě [10]

Kohonenova vrstva je tvořena neurony y , jež jsou uspořádané do topologické struktury - nejčastěji to bývá dvojrozměrná mřížka (Obrázek 16). Topologická struktura určuje, které neurony se sebou sousedí, což je potřeba znát v procesu učení. Každému neuronu zde přísluší vektor vah o stejné dimenzi, jakou má vstupní vrstva. Tato vrstva je propojena se vstupní vrstvou. Způsob, jakým se zjišťuje pozice primární odezvy v SOM na aktuální podnět, se jmenuje soutěžení. Výsledkem soutěžení v každém kroku (po předložení konkrétního vstupu) je vítězný neuron, který nejvíc reaguje na daný vstup x .

Kohonenův algoritmus samoorganizace [10]:

Vstup: trénovací množina $x_i, i = 1, 2, \dots, N$.

Výstup: vektory vah $w_j, j = 1, 2, \dots, K$.

1. inicializuj náhodně $w_j, i = 0$,
2. vezmi x_i a nalezni i^* -tý neuron,
3. $\|x_i - w_{i^*}\| = \min_j \|x_i - w_j\|$,
4. Uprav všechny váhové vektory (pro všechna j): $w_i = w_j + \alpha \cdot h_{i^*j}(x_i - w_j)$, kde α je koeficient útlumu a h_{i^*j} je funkce okolí, $i = i + 1$,
5. opakuj kroky 2 až 4 do splnění kritéria zastavení.

Kde:

x - vstup; i^* - index vítězného neuronu y_{i^*} ; i - index neuronů y ; j - index vektorů vah.

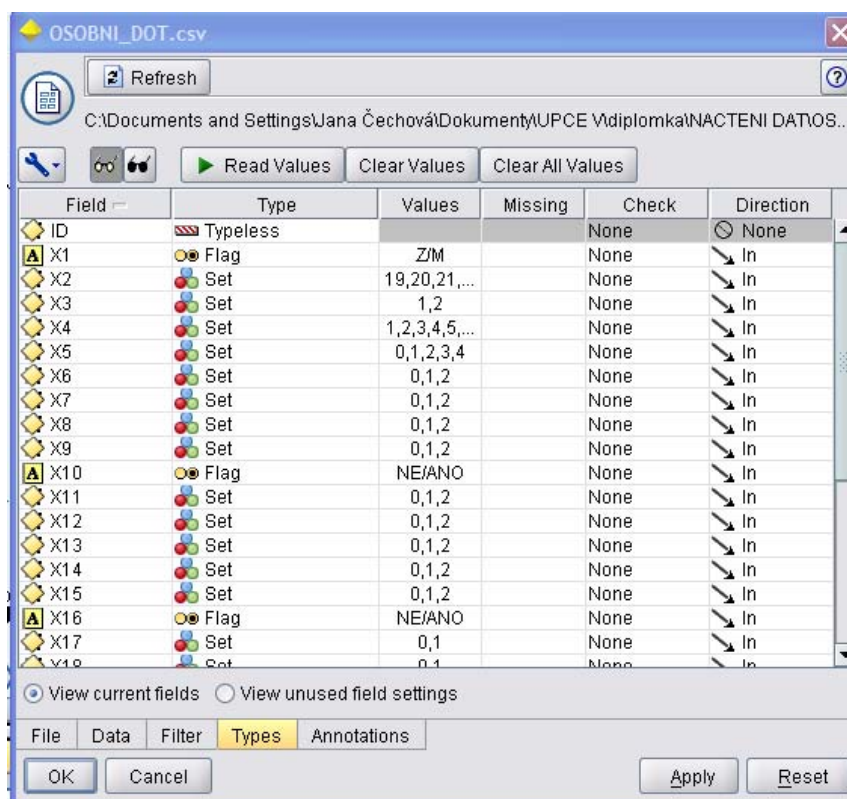
6. Klasifikace studentů

Tato kapitola se věnuje přímému zpracování připravených dat. Prvním krokem je načtení dat do prostředí SPSS Clementine 10.1 a jejich potřebné úpravy. Dále následují konkrétní kroky k získání výsledku, tedy samotné modelování.

6.1 Načtení dat

Před zahájením procesu modelování jsou načtena data v použitelném formátu - ve formě tabulky. Matice O a S byly vytvořeny v prostředí MS Excel. Pro načtení dat do SPSS Clementine je nutné uložit tabulky ve formátu CSV. Jedná se o soubor s proměnnou délkou záznamu, kde je jako oddělovač polí použit středník [1].

Po načtení dat jsou nastaveny parametry: cesta a název souboru, oddělovač (středník). Následuje specifikace typu dat v jednotlivých polích (Obrázek 17).



Obrázek 17 – Uzel Var. File, karta Types po načtení dat [vlastní]

Pro úvodní analýzu dat je proveden datový audit (Data Audit). Umožňuje analyzovat data ze vstupního souboru a zobrazit výsledky této analýzy. Jedná se o následující charakteristiky dat z jednotlivých polí – název pole, zobrazení dat ve formě grafu odpovídajícího typu dat v poli, typ pole, minimální, maximální, střední hodnota, standardní odchylku a nesouměrnost dat. Dále je zde uveden počet jedinečných hodnot a počet správných hodnot [11]. Obrázek 18 dokumentuje výsledek tohoto uzlu v případě zdrojové matice S.

Field	Graph	Type	Min	Max	Mean	Std. Dev	Skewness	Unique	Valid
X31		Flag	--	--	--	--	--	1	495
X32		Flag	1	2	--	--	--	2	495
X33		Ordered Set	1	5	--	--	--	5	495
X34		Set	--	--	--	--	--	3	495
X35		Set	--	--	--	--	--	10	495
X36		Set	--	--	--	--	--	3	495
X37		Set	1	4	--	--	--	4	495
X38		Flag	0	1	--	--	--	2	495

Obrázek 18 – Uzel Data Audit ze souboru Studijni_program [vlastní]

Jak je vidět na Obrázku 18, proměnné X_{31} až X_{38} jsou kompletní, neobsahují žádnou prázdnou hodnotu (chybějící hodnotu), kterou je třeba nahradit. Možnosti náhrady chybějících hodnot jsou popsány v následující kapitole.

6.1.1 Nahrazení chybějících hodnot

V reálných úlohách dobývání znalostí je možné dostat se do situace, kdy chybí údaje o některých objektech. Problém práce s chybějícími hodnotami lze řešit v kroku předzpracování. Na náhradu chybějících hodnot existuje více způsobů [1]:

- ignorovat objekt s nějakou chybějící hodnotou,
- nahradit chybějící hodnotu novou hodnotou „nevím“,
- nahradit chybějící hodnotu některou z existujících hodnot atributu a to:
 - nejčtetnější hodnotou,
 - proporcionálním podílem všech hodnot,
 - libovolnou hodnotou.

V případě této práce již došlo v průběhu předzpracování dat k nahrazení chybějících hodnot v datové matici O a S.

6.2 Rozhodovací stromy

Jak již bylo uvedeno v kapitole 5.3.1, jsou vytvářeny modely pro klasifikaci studentů na základě cílových proměnných. Cílové proměnné jsou stanoveny jako výstupy pro klasifikaci.

Z datové matice jsou vždy vybrány vhodné vstupní atributy, u kterých jsou rozhodovací stromy analyzovány závislosti. Výsledkem je nalezení stromu konzistentního s trénovacími daty, přitom se dává přednost menším, jednodušším stromům [1].

Po načtení jsou data rozdělena na trénovací a testovací množinu uzlem Partition. Tento uzel se používá pro jednoduché ověření kvality získaných modelů. Nejprve je zkoumáno, v jakém poměru budou data dělena na trénovací a testovací. Nakonec je jako nejvhodnější varianta u všech uzlů zvolena kombinace 75:25 (trénovací : testovací). Po nastavení vstupních proměnných a výstupní proměnné dochází k určování závislostí v datech díky uzlům CHAID, QUEST, C&RT a C5.0. Výstupy těchto modelů jsou vyhodnocovány pomocí uzlu Analysis a grafu Evaluation. Uzel Analysis slouží pro kvalitativní analýzu predikčních modelů, dochází zde k vzájemnému porovnání modelů. Uzel Evaluation poskytuje evaluační grafy pro hodnocení kvality a vzájemné komparace predikčních modelů. Jeho výsledkem jsou křivky jednotlivých modelů pro tvorbu RS.

6.2.1 Osobní dotazování

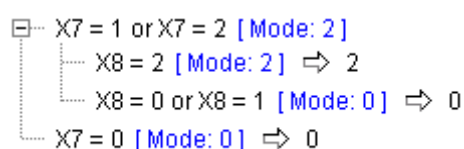
V této části jsou modelovány výstupy poukazující na osobní záležitosti každého studenta jako je bydlení, stravování, zdroje financí či volný čas.

6.2.1.1 Spokojenost s cenou kolejí

Protože expertní odhad podává velmi jednoduchý výsledek, jsou výstupní atributy pro klasifikaci této proměnné vygenerovány pomocí čtyř metod - expertní odhad, shluková analýza, rozhodovací stromy a uzlem Feature Selection. Výsledky jsou poté porovnány a klasifikace je provedena nejvhodnějším způsobem.

Expertní odhad

Expertním odhadem byly stanoveny vstupní atributy: pohlaví (x_1), věk (x_2), rodinný stav (x_3), bydliště v Pardubicích (x_5), cena odpovídá službám kolejí (x_7), spokojenost s přidělováním kolejí (x_8), souhlas s propojením účtu kolejí a menzy (x_9), stravování v menze (x_{10}), platba studia (x_{20}), brigáda (x_{21}), vlastník automobilu (x_{24}), vyhovující parkování v okolí Univerzity Pardubice (UPCE) (x_{25}), spokojenost s okolím UPCE (x_{26}). Po provedení analýzy pomocí RS byl vyhodnocen jako nejvhodnější model pro tvorbu RS uzel QUEST. Rozhodovací pravidla určena tímto stromem vypadají následovně:



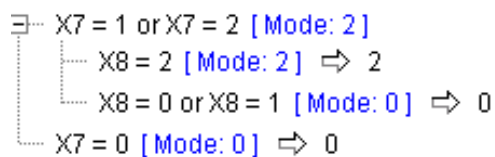
Obrázek 19 – Rozhodovací pravidla uzlu QUEST (x_6) [vlastní]

Výsledkem Expertního odhadu je jednoduchý strom, který vykazuje závislosti především na proměnné x_8 (spokojenost s přidělovacím systémem na kolejích).

Rozhodovací stromy

Vstupem pro RS jsou nejprve zvoleny všechny atributy. Výsledek není směrodatný pro konečnou klasifikaci. Atributy vyhodnocené RS jako důležité slouží jako zdroj vstupních atributů pro další modelování. Vybrán byl dle kvality klasifikace uzel QUEST. Na základě tohoto modelu je dále pracováno s atributy: kraj (x_4), bydliště v Pardubicích (x_5), cena odpovídá službám kolejí (x_7), spokojenost s přidělováním kolejí (x_8), souhlas s propojením účtu kolejí a menzy (x_9).

Konečným výsledkem RS jsou následující rozhodovací pravidla opět uzlu QUEST (Obrázek 20).



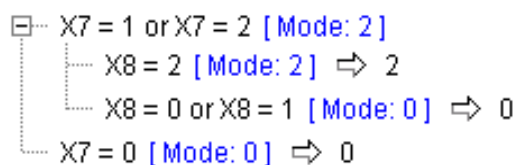
Obrázek 20 – Rozhodovací pravidla uzlu QUEST (x_6) [vlastní]

Výsledek je tedy stejný jako u předchozí metody.

Shluková analýza

Podle výsledků shlukové analýzy (Kohonenovy mapy), jsou určeny důležité atributy pro výzkum pomocí RS. Vstupními atributy byly tentokrát určeny: stravování v menze (x_{10}), spokojenost s cenou jídel (x_{11}), cena odpovídá kvalitě jídel (x_{12}), spokojenost s otvírací dobou (x_{13}), spokojenost se systémem volného výběru jídel (x_{14}), doplňkový sortiment menzy (x_{15}), katedra tělesné výchovy (TV) (x_{16}), povinná TV (x_{17}), výcviková forma (x_{19}), věk (x_2), vlastník automobilu (x_{24}), vyhovující parkování v okolí UPCE (x_{25}), spokojenost s okolím UPCE (x_{26}), kraj (x_4), bydliště v Pardubicích (x_5), spokojenost s cenou kolejí (x_6), cena odpovídá službám kolejí (x_7), spokojenost s přidělováním kolejí (x_8), souhlas s propojením účtu kolejí a menzy (x_9).

Výsledky provedených algoritmů jsou nejkvalitnější u uzlu QUEST. Obrázek 21 ukazuje výsledná rozhodovací pravidla.

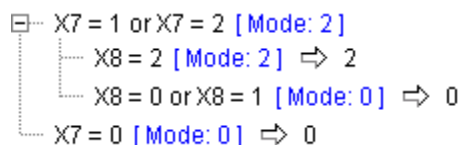


Obrázek 21 - Rozhodovací pravidla uzlu QUEST(x_6) [vlastní]

Uzel Feature Selection

Důležitou funkcí algoritmu Feature selection je použití na identifikaci polí, která jsou velmi důležitá pro danou analýzu. Je zadáno jedno cílové pole (Out) a zvolený počet prediktorů, jež je třeba prověřit nebo vyhodnotit relativně k cílovému poli. Jak cílové pole, tak prediktory mohou být kategorické i rozsahové [11].

Po vyhodnocení důležitých atributů uzlem Feature Selection (cena odpovídá službám kolejí (x_7), spokojenost s přidělováním kolejí (x_8), souhlas s propojením účtu kolejí a menzy (x_9), kraj (x_4), bydliště v Pardubicích (x_5), vyhovující parkování v okolí UPCE (x_{25}), doplňkový sortiment menzy (x_{15}), výcviková forma TV (x_{19}), oblíbené kluby v Pardubicích (x_{30}), spokojenost s okolím UPCE (x_{26})), jsou RS vygenerována následující rozhodovací pravidla (Obrázek 22). Uzel QUEST je vyhodnocen jako nejkvalitnější pro klasifikaci a jeho výsledek je opět totožný s předchozími metodami.



Obrázek 22 - Rozhodovací pravidla uzlu QUEST(x_6) [vlastní]

Vyhodnocení klasifikace

Porovnáním výstupů jednotlivých metod lze konstatovat, že fakt, že student je spokojen s cenou kolejí, je ovlivněn vždy pouze tím, odpovídá-li tato cena kvalitě služeb poskytovaných na koleji a zároveň je-li spokojen s přidělovacím systémem kolejí.

6.2.1.2 Spokojenost s cenou jídla

Klasifikace studentů na základě výstupního atributu spokojenost s cenou jídla v menze (x_{11}) je opět provedena čtyřmi metodami. Expertní odhad, shluková analýza, rozhodovací stromy a uzel Feature Selection vygenerují důležité atributy, které jsou následně použity pro modelování RS.

Expertní odhad

Expertním odhadem byly stanoveny na základě logického výběru vstupní atributy pro dané modely RS (pohlaví (x_1), věk (x_2), bydliště v Pardubicích (x_5), spokojenost s cenou kolejí (x_6), cena odpovídá kvalitě jídel (x_{12}), spokojenost s otvírací dobou (x_{13}), spokojenost se systémem volného výběru jídel (x_{14}), doplňkový sortiment menzy (x_{15}), katedra TV (x_{16}), platba studia (x_{20}), brigáda (x_{21})). Podle kvality výstupu je vybrán nejvhodnější uzel pro klasifikaci - CHAID. Porovnání jednotlivých kvality jednotlivých RS ukazuje Obrázek 23.

Results for output field X11

Individual Models

Comparing \$C-X11 with X11

'Partition'	1_Training		2_Testing	
Correct	317	87,57%	109	81,95%
Wrong	45	12,43%	24	18,05%
Total	362		133	

Comparing \$R-X11 with X11

'Partition'	1_Training		2_Testing	
Correct	317	87,57%	109	81,95%
Wrong	45	12,43%	24	18,05%
Total	362		133	

Comparing \$R1-X11 with X11

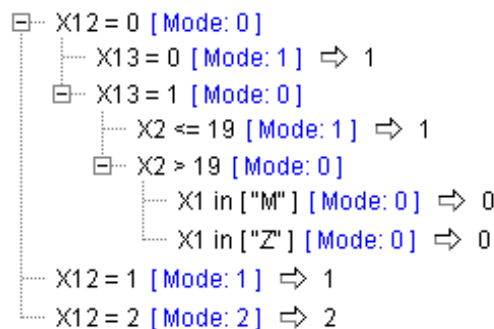
'Partition'	1_Training		2_Testing	
Correct	317	87,57%	109	81,95%
Wrong	45	12,43%	24	18,05%
Total	362		133	

Comparing \$R2-X11 with X11

'Partition'	1_Training		2_Testing	
Correct	289	74,31%	90	67,67%
Wrong	93	25,69%	43	32,33%
Total	362		133	

Obrázek 23 – Vyhodnocení kvality výstupů RS (x_{11}) [vlastní]

Model klasifikoval studenty podle následujících rozhodovacích pravidel:



Obrázek 24 - Rozhodovací pravidla uzlu CHAID (x_{11}) [vlastní]

Rozhodovací stromy

Vstupem pro RS jsou nejprve zvoleny všechny atributy matice O. Protože tento výsledek není pro klasifikaci příliš vhodný, jsou vybrány podle nejvhodnějšího algoritmu pouze atributy, vyhodnocené daným uzlem jako nejdůležitější (pohlaví (x_1), spokojenost s cenou kolejí (x_6), cena odpovídá kvalitě jídel (x_{12}), spokojenost s otvírací dobou (x_{13}), katedra TV (x_{16}), povinná TV (x_{17}), volitelná TV (x_{18}), brigáda (x_{21}), spokojenost s okolím UPCE (x_{26}), volný čas – kulturní akce (x_{29})). Tyto atributy jsou dále použity jako vstupní pro opakovanou analýzu.

Results for output field X11

Individual Models

Comparing \$C-X11 with X11

'Partition'	1_Training		2_Testing	
Correct	330	91,16%	106	79,7%
Wrong	32	8,84%	27	20,3%
Total	362		133	

Comparing \$R-X11 with X11

'Partition'	1_Training		2_Testing	
Correct	330	91,16%	105	78,95%
Wrong	32	8,84%	28	21,05%
Total	362		133	

Comparing \$R1-X11 with X11

'Partition'	1_Training		2_Testing	
Correct	318	87,85%	115	86,47%
Wrong	44	12,15%	18	13,53%
Total	362		133	

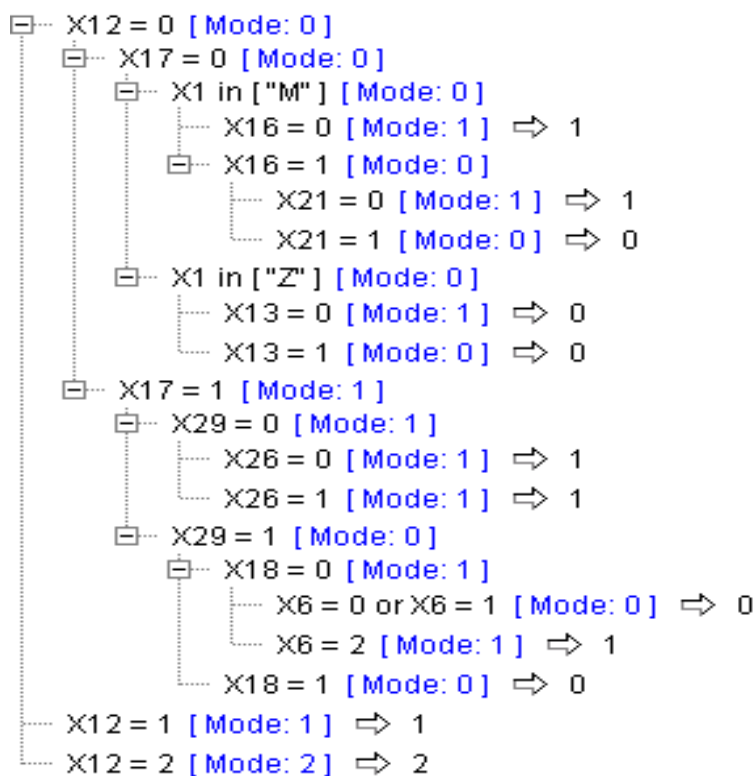
Comparing \$R2-X11 with X11

'Partition'	1_Training		2_Testing	
Correct	273	75,41%	83	62,41%
Wrong	89	24,59%	50	37,59%
Total	362		133	

Obrázek 25 - Vyhodnocení kvality výstupů RS (x_{11}) [vlastní]

Po použití těchto vstupních atributů je z výstupů jednotlivých uzlů vybrán uzel CHAID.

Rozhodovací pravidla tohoto uzlu vypadají následovně:



Obrázek 26 - Rozhodovací pravidla uzlu CHAID (x_{11}) [vlastní]

Shluková analýza

Podle výsledků shlukové analýzy (Kohonenovy mapy) jsou opět vyhodnoceny důležité atributy pro výzkum pomocí RS (stravování v menze (x_{10}), spokojenost s cenou jídel (x_{11}), cena odpovídá kvalitě jídel (x_{12}), spokojenost s otvírací dobou (x_{13}), spokojenost se systémem volného výběru jídel (x_{14}), doplňkový sortiment menzy (x_{15}), katedra TV (x_{16}), povinná TV (x_{17}), výcviková forma (x_{19}), věk (x_2), vlastník automobilu (x_{24}), vyhovující parkování v okolí UPCE (x_{25}), spokojenost s okolím UPCE (x_{26}), kraj (x_4), bydliště v Pardubicích (x_5), spokojenost s cenou kolejí (x_6), cena odpovídá službám kolejí (x_7), spokojenost s přidělováním kolejí (x_8), souhlas s propojením účtu kolejí a menzy (x_9)).

Results for output field X11

Individual Models

Comparing \$C-X11 with X11

Partition'	1_Training		2_Testing	
Correct	317	87,57%	111	83,46%
Wrong	45	12,43%	22	16,54%
Total	362		133	

Comparing \$R-X11 with X11

Partition'	1_Training		2_Testing	
Correct	310	85,64%	106	79,7%
Wrong	52	14,36%	27	20,3%
Total	362		133	

Comparing \$R1-X11 with X11

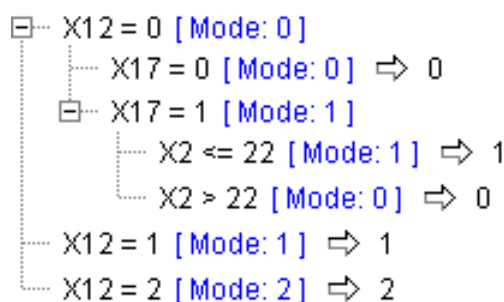
Partition'	1_Training		2_Testing	
Correct	317	87,57%	111	83,46%
Wrong	45	12,43%	22	16,54%
Total	362		133	

Comparing \$R2-X11 with X11

Partition'	1_Training		2_Testing	
Correct	269	74,31%	90	67,67%
Wrong	93	25,69%	43	32,33%
Total	362		133	

Obrázek 27 - Vyhodnocení kvality výstupů RS (x_{11}) [vlastní]

Po provedení algoritmu RS s těmito vstupními atributy je vyhodnocen jako nejvhodnější model C5.0 (Obrázek 27). Výsledek modelu je interpretován následujícími rozhodovacími pravidly uvedenými na Obrázku 28.



Obrázek 28 - Rozhodovací pravidla uzlu C5.0 (x_{11}) [vlastní]

Uzel Feature Selection

Na základě tohoto uzlu jsou vyhodnoceny jako důležité následující vstupy: cena jídel = kvalita jídel (x_{12}), spokojenost s otvírací dobou menzy (x_{13}), doplňkový sortiment menzy (x_{15}), spokojenost se systémem volného výběru jídel (x_{14}), stravování v menze (x_{10}), věk (x_2), povinná TV (x_{17}), volitelná TV (x_{18}), čas strávený na brigádě (x_{22}), bydliště v Pardubicích (x_5), spokojenost s přidělováním kolejí (x_8), spokojenost s cenou kolejí (x_6). Po jejich použití je výstup modelu rozhodovacích stromů uveden na Obrázku 29.

Results for output field X11

Individual Models

Comparing \$C-X11 with X11

Partition	1_Training		2_Testing	
Correct	322	88,95%	107	80,45%
Wrong	40	11,05%	26	19,55%
Total	362		133	

Comparing \$R-X11 with X11

Partition	1_Training		2_Testing	
Correct	317	87,57%	115	86,47%
Wrong	45	12,43%	18	13,53%
Total	362		133	

Comparing \$R1-X11 with X11

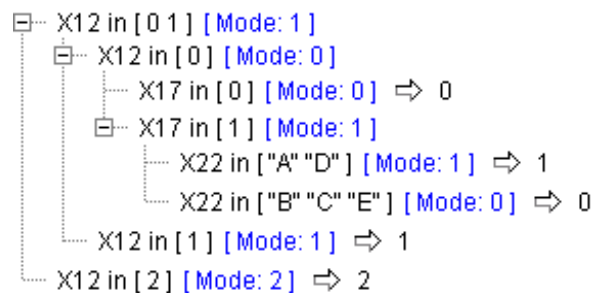
Partition	1_Training		2_Testing	
Correct	317	87,57%	115	86,47%
Wrong	45	12,43%	18	13,53%
Total	362		133	

Comparing \$R2-X11 with X11

Partition	1_Training		2_Testing	
Correct	273	75,41%	83	62,41%
Wrong	89	24,59%	50	37,59%
Total	362		133	

Obrázek 29 - Vyhodnocení kvality výstupů RS (x_{11}) [vlastní]

Nejkvalitnější výsledky podává uzel C&RT. Rozhodovací pravidla vypadají následovně:



Obrázek 30 - Rozhodovací pravidla uzlu C&RT (x_{11}) [vlastní]

Vyhodnocení klasifikace

Klasifikace studentů podle cílové proměnné spokojenost studentů s cenou jídla v menze je podle výše uvedených metod ovlivněna především atributy cena jídel = kvalita jídel (x_{12}) a povinná TV (x_{17}). Zajímavá je závislost na proměnné x_{17} (využití katedry tělesné výchovy – povinný TV). V případě expertních odhadů tento atribut nebyl brán v úvahu, ale ostatní metody ho vyhodnotili jako druhý nejdůležitější.

Konečná klasifikace studentů je provedena podle nejlepších výsledků chybných hodnot (minimálních hodnot) v trénovací a testovací množině dat.

Tabulka 14 – Přesnost klasifikace [vlastní]

Použité metody	Počet chybných hodnot	
	Trénovací množina	Testovací množina
Expertní odhad	45	24
Rozhodovací stromy	32	27
Shluková analýza	45	22
Feature selection	45	18

Podle Tabulky č. 14 nelze jednoznačně říci, která metoda má nejlepší výsledky. V úvahu připadají dvě zkoumané metody: RS a Feature Selection. Po nahlédnutí na poměr dat trénovacích a testovacích je možné porovnat tyto dvě metody (Tabulka č. 14).

Tabulka 15 – Konečné porovnání metod [vlastní]

Použité metody	Počet správně zařazených záznamů[%]	
	Trénovací množina	Testovací množina
Rozhodovací stromy	91,16	79,7
Feature Selection	87,57	86,47

Protože odchylka trénovacích a testovacích dat je menší u modelu, kde je použit uzel Feature Selection, je pro klasifikaci studentů podle cílové proměnné Spokojenost s cenou jídla použit RS tohoto modelu (Příloha č. 8).

Atribut spokojenost s cenou může nabývat třech hodnot – spokojen, nespokojen a neodpověděl. Jak je uvedeno v počáteční analýze 338 studentů využívá služeb menzy. Klasifikace je zaměřena na studenty, kteří jsou spokojeni s menzou, což je 151 studentů a 93 studentů je nespokojeno. Prvním atributem pro klasifikaci je na základě algoritmu označeno to, odpovídá-li cena kvalitě jídla. Ze spokojených studentů 101 souhlasí s tím,

že cena odpovídá kvalitě. Naopak třetina spokojených studentů odmítá připustit, že cena odpovídá kvalitě. Zajímavou spojitost našel algoritmus s atributem povinná tělesná výchova. Studenti nespokojení s kvalitou jídla jsou klasifikováni podle toho, navštěvují-li povinnou TV. 21 studentů TV nenavštěvuje a 29 ano. Studenti navštěvující povinnou TV jsou dále klasifikováni podle doby, jakou pracují. 11 studentů spokojených s cenou jídla, nespokojených s kvalitou a navštěvujících povinnou TV pracují za týden 1-5 hodin, nebo 20 a více. To je kombinace minimální a maximální nabízené doby pro odpověď. Ostatních 18 studentů chodí na brigádu 5 - 10, 10 - 20 hodin týdně, nebo neodpověděli.

6.2.1.3 Financování

Cílová proměnná financování je nejprve vygenerována uzlem Derive na základě atributu platba studia (x_{20}). Pokud se platby studia studentů účastní rodiče, proměnná nabývá hodnot „Nesamostatní“. V případě, že si student platí studium sám, nebo dostává stipendium, nabývá proměnná hodnoty „Samostatní“. Ostatní studenti, kteří neodpověděli, zastupují hodnotu této proměnné „Neodpověděl“. Pro následné modelování jsou hodnoty převedeny na kódy 0, 1, 2.

Vstupní proměnné jsou definovány následovně: pohlaví (x_1), věk (x_2), bydliště v Pardubicích (x_5), brigáda (x_{21}), vlastník živnostenského listu (x_{23}), vlastník automobilu (x_{24}).

Výstupní porovnání kvality dat lze v tomto případě provést pouze uzlem Analysis. Jak je vidět na Obrázku 31, všechny uzly mají stejný počet případů, kde se klasifikátor neshoduje s učitelem. Nelze určit minimální či maximální hodnotu, protože všechny výstupní hodnoty jsou naprosto totožné.

Results for output field Financ_ciselne

Individual Models

Comparing \$C-Financ_ciselne with Financ_ciselne

'Partition'	1_Training		2_Testing	
Correct	314	86,74%	118	88,72%
Wrong	48	13,26%	15	11,28%
Total	362		133	

Comparing \$R-Financ_ciselne with Financ_ciselne

'Partition'	1_Training		2_Testing	
Correct	314	86,74%	118	88,72%
Wrong	48	13,26%	15	11,28%
Total	362		133	

Comparing \$R1-Financ_ciselne with Financ_ciselne

'Partition'	1_Training		2_Testing	
Correct	314	86,74%	118	88,72%
Wrong	48	13,26%	15	11,28%
Total	362		133	

Comparing \$R2-Financ_ciselne with Financ_ciselne

'Partition'	1_Training		2_Testing	
Correct	314	86,74%	118	88,72%
Wrong	48	13,26%	15	11,28%
Total	362		133	

Obrázek 31 - Vyhodnocení kvality výstupů RS (Financování) [vlastní]

Po nadefinování vstupů byl jako jediný použitelný strom vygenerován výsledek uzlu CHAID (Příloha č. 9). Tento uzel používá pro větvení 3 atributy (pohlaví, brigáda, vlastník živnostenského listu). Naprostá většina studentů z množiny testovacích dat byla klasifikována jako „Nesamostatní“ (0) - 314 studentů. Ostatní studenti jsou „Samostatní“ (1) - 46 studentů nebo na danou otázku neodpověděli – hodnota „ Neodpověděl“ (2) - 2 studenti. Ze studentů, kteří jsou nesamostatní, 171 studentů nechodí na brigádu a 114 na brigádu chodí. Ti studenti, kteří nechodí na brigádu a jsou nesamostatní ve většině případů (169 studentů) nevlastní živnostenský list. Vlastníkem živnostenského listu v této kategorii jsou pouze dva studenti. Nesamostatní studenti, kteří nechodí na brigádu a nemají živnostenský lis, jsou zastoupeni ve většině ženami (136 studentů). Mužů je v této kategorii pouze 33.

6.2.2 Studijní program

Matrice S obsahuje odpovědi na otázky zaměřené na studium a oblasti s ním spojené. Kromě základních analýz o počtech studentů jednotlivých programů a oborů je zajímavé zkoumat v této matici i ostatní závislosti mezi studenty. V následující části budou zkoumány závislosti u opakování předmětů – opakování matematických předmětů, opakování předmětů informatiky, využití možnosti studia v zahraničí a stanovení cílů studentů.

6.2.2.1 Opakování předmětů

Jako cílová (vysvětlovaná) proměnná je určen nejdříve atribut Opakování předmětů (x_{39}). Vstupem pro tento model jsou zvoleny následující vysvětlující proměnné: typ studia (x_{32}), ročník studia (x_{33}), studijní program (x_{34}), přijetí ke studiu (x_{36}), typ střední školy (x_{37}), prospěchové stipendium (x_{38}), preferování předtermínů (x_{49}) a zkouškové termíny (x_{50}).

Po vygenerování výsledků je pomocí porovnání kvality výstupů (Obrázek 32) vybrán uzel s nejhodnějším poměrem správných a chybných dat.

Results for output field X39

Individual Models

Comparing \$C-X39 with X39

'Partition'	1_Training		2_Testing	
Correct	288	79,56%	102	76,69%
Wrong	74	20,44%	31	23,31%
Total	362		133	

Comparing \$R-X39 with X39

'Partition'	1_Training		2_Testing	
Correct	285	78,73%	107	80,45%
Wrong	77	21,27%	26	19,55%
Total	362		133	

Comparing \$R1-X39 with X39

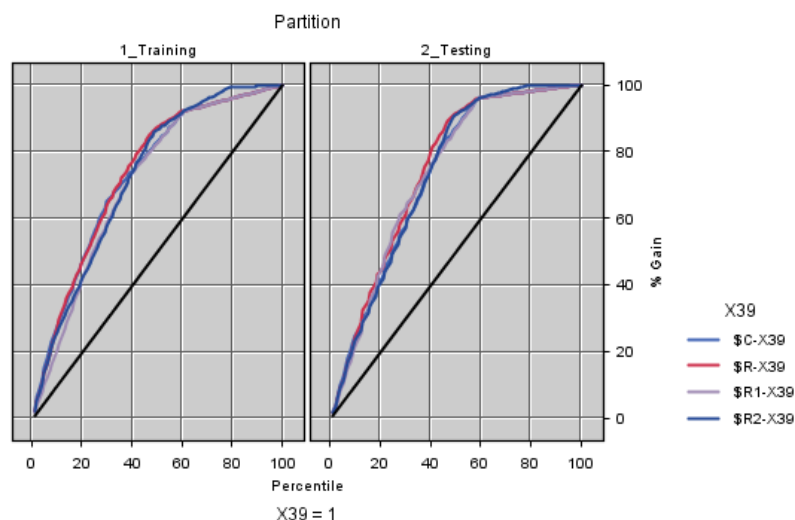
'Partition'	1_Training		2_Testing	
Correct	274	75,69%	107	80,45%
Wrong	88	24,31%	26	19,55%
Total	362		133	

Comparing \$R2-X39 with X39

'Partition'	1_Training		2_Testing	
Correct	271	74,86%	94	70,68%
Wrong	91	25,14%	39	29,32%
Total	362		133	

Obrázek 32 – Vyhodnocení kvality výstupů RS (x_{39}) [vlastní]

Vzhledem k počtu případů, kde se klasifikátor neshoduje s učitelem je vybrán RS, který je výsledkem uzlu C&RT. Počet případů je zde minimální ze všech RS. Dalším vodítkem pro výběr RS je výstup uzlu Evaluation. Porovnání modelů pomocí tohoto uzlu vypadá následovně (Obrázek 33).



Obrázek 33 – Výstup uzlu Evaluation (x_{39}) [vlastní]

Podle obrázku je zřejmé, že všechny modely klasifikují podobně. S naprostou jistotou nelze stanovit uzel s nejvyšší kvalitou výstupu. Maximální kvality klasifikace je u všech modelů dosaženo na přibližně 50 percentilech.

Rozhodovací strom v grafické podobě (Příloha č. 10) který má největší vypovídací schopnost je výsledkem uzlu C5.0. RS má největší logickou návaznost s dělením podle pěti atributů. Prvním atributem, podle kterého se strom dělí je ročník studia, druhým preferované zkouškové termíny, třetím je typ studia (bakalářský x navazující magisterský), čtvrtým je studijní program (HPS, EM, SII) a posledním atributem je typ střední školy studenta.

V testovací množině je 362 studentů. Většina z nich, 249 studentů, neopakuje žádný předmět. Ze zbylých 113, kteří opakují minimálně jeden předmět, je 9 studentů z prvního ročníku a 104 studentů tvoří kategorii druhého a třetího ročníku, což je naprosto logické (studenti prvního ročníku mají všechny předměty poprvé). Řádné termíny zkoušek využívá naprostá většina studentů (100), zbylí 4 studenti preferují mimořádné termíny zkoušek (např. září). Ze stovky studentů, kteří opakují a preferují řádné termíny zkoušek je 20 studentů navazujícího magisterského studia a 80 bakalářů. Z těchto bakalářů, kteří opakují nějaké předměty a preferují řádné zkouškové termíny je 57 studentů v studijním programu HPS, 14 EM a 9 SII. Posledním atributem pro klasifikaci je rozdělení studentů programu HPS podle střední školy. Z RS je patrné že nejvíce studentů (40) opakujících nějaké předměty je z obchodní akademie, méně z gymnázií (10 studentů) a polední zastoupenou skupinou jsou studenti z jiných středních škol (7).

6.2.2.2 Opakování matematických předmětů

Protože největší procento ze studentů, kteří opakují, jsou studenti opakující matematické a inženýrské předměty, je zajímavé studovat závislosti těchto výstupních proměnných na vstupních atributech.

Cílová proměnná je opakování matematiky (x_{41}). Jako vstup pro model jsou zvoleny následující proměnné: typ studia (x_{32}), ročník studia (x_{33}), studijní program (x_{34}), přijetí ke studiu (x_{36}), typ střední školy (x_{37}), prospěchové stipendium (x_{38}), opakování předmětů (x_{39}), preferování předtermínů (x_{49}) a zkouškové termíny (x_{50}).

Results for output field X41

Individual Models

Comparing \$C-X41 with X41

'Partition'	1_Training		2_Testing	
Correct	328	90,61%	110	82,71%
Wrong	34	9,39%	23	17,29%
Total	362		133	

Comparing \$R-X41 with X41

'Partition'	1_Training		2_Testing	
Correct	330	91,16%	107	80,45%
Wrong	32	8,84%	26	19,55%
Total	362		133	

Comparing \$R1-X41 with X41

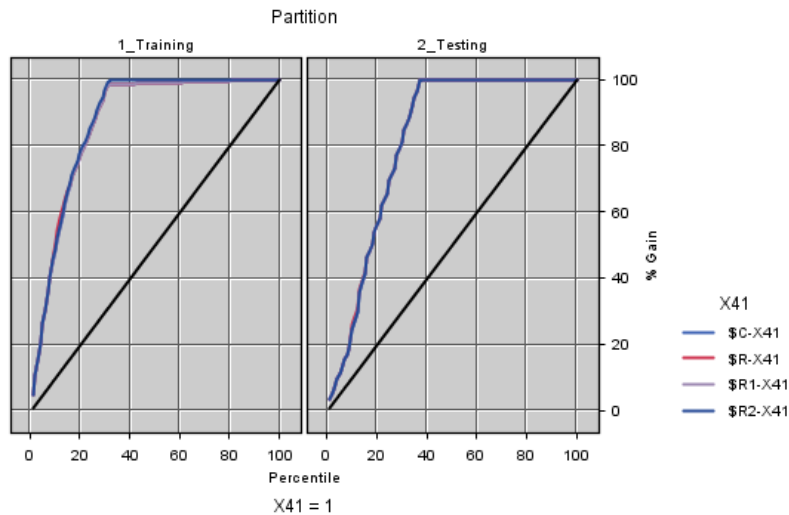
'Partition'	1_Training		2_Testing	
Correct	328	90,61%	110	82,71%
Wrong	34	9,39%	23	17,29%
Total	362		133	

Comparing \$R2-X41 with X41

'Partition'	1_Training		2_Testing	
Correct	329	90,88%	110	82,71%
Wrong	33	9,12%	23	17,29%
Total	362		133	

Obrázek 34 – Vyhodnocení kvality výstupů RS (x_{41}) [vlastní]

Podle výstupu uzlu Analysis (Obrázek 34) je kvalita dat ohodnocena velmi podobně. Podle konkrétních hodnot je jako nejvhodnější výstup vyhodnocen výstup uzlu CHAID (počet případů, kde se klasifikátor neshoduje s učitelem je minimální). Výstup grafu Evaluation nyní ukazuje, že křivky u všech modelů jsou naprosto totožné. Pro volbu modelu zde není tento výstup směrodatný. Ukazuje, že maximální kvalita klasifikace je na hranici 40 percentilu.



Obrázek 35 – Výstup uzlu Evaluation (x_{41}) [vlastní]

Grafický výstup RS (Příloha č. 11) zde měl největší a nejlogičtější vypovídací schopnost u uzlu CHAID. Ten zahrnul do větvení atributy opakování předmětu, ročník studia, typ studia a oproti ostatním větvil strom i podle atributu prospěchové stipendium.

Testovací množina pracuje opět s 362 záznamy. Matematiku neopakuje 299 studentů a 63 ji opakuje. Z počtu 299 studentů, kteří neopakují matematiku je 248 studentů, kteří neopakují žádný předmět a 51 studentů, kteří neopakují matematiku, ale jiný předmět opakují. Z klasifikovaných, kteří neopakují ani jeden předmět je 12 studentů, kteří pobírají prospěchové stipendium. Celá tato skupina dvanácti studentů je z prvního nebo druhého ročníku. Naopak studenti, kteří opakují matematické předměty (63) jsou ve většině z bakalářského studia (40 studentů). Z těchto čtyřiceti studentů je 5 studentů v prvním ročníku, 19 v prvním či druhém ročníku a 16 v ročníku třetím.

6.2.2.3 Opakování informatických předmětů

Jako výstup je v tomto modelu stanovena proměnná opakování informatiky (x_{43}). Vstupní proměnné jsou stejné jako u předchozího modelu (opakování matematických předmětů).

Výstupy modelování jsou hodnoceny následujícím způsobem (Obrázek 36).

Results for output field X43

Individual Models

Comparing \$C-X43 with X43

Partition	1_Training	2_Testing
Correct	337 93,09%	117 87,97%
Wrong	25 6,91%	16 12,03%
Total	362	133

Comparing \$R-X43 with X43

Partition	1_Training	2_Testing
Correct	338 93,37%	116 87,22%
Wrong	24 6,63%	17 12,78%
Total	362	133

Comparing \$R1-X43 with X43

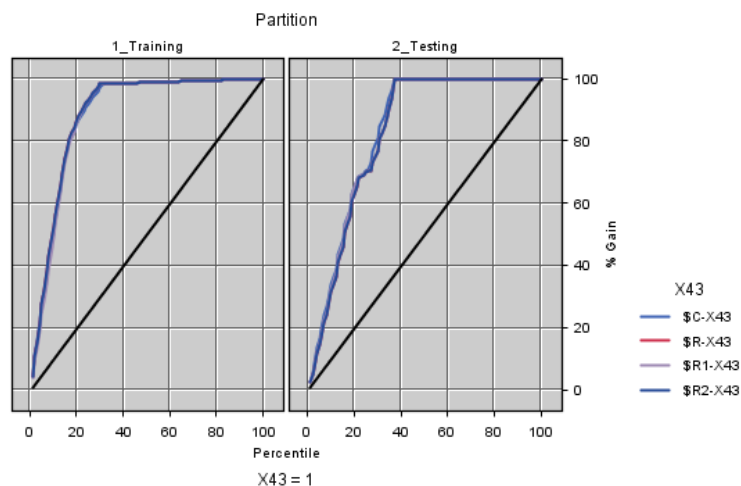
Partition	1_Training	2_Testing
Correct	336 92,82%	117 87,97%
Wrong	26 7,18%	16 12,03%
Total	362	133

Comparing \$R2-X43 with X43

Partition	1_Training	2_Testing
Correct	338 93,37%	116 87,22%
Wrong	24 6,63%	17 12,78%
Total	362	133

Obrázek 36 – Vyhodnocení kvality výstupů RS (x_{43}) [vlastní]

Kvalita jednotlivých uzlů je opět téměř totožná. S malou odlišností je zde možné z minimálního počtu případů, kde se klasifikátor neshoduje s učitelem určit jako nejvhodnější uzel C5.0. Měření kvality výstupu grafem Evaluation opět nemá žádný význam. Křivky optimálních modelů jsou takřka totožné (Obrázek 37). Maximální kvality klasifikace je jako v předchozím případě (opakování matematiky) dosaženo u všech modelů na hranici 40 percentilu.



Obrázek 37 – Výstup uzlu Evaluation (x_{43}) [vlastní]

Grafická podoba stromu uzlu C5.0 je uvedena v Příloze č. 12. Jako v předchozích modelech používá testovací množina 362 záznamů. 304 studentů neopakuje informatické předměty a 58 zbylých studentů tyto předměty opakuje. Téměř všichni z 58 studentů opakujících informatiku jsou v bakalářském programu studia (53 studentů).

Zbylí studenti jsou v navazujícím studijním programu. Největší část „bakalářů“ je z programu HPS (38 studentů), z prvního či druhého ročníku je zastoupeno 32 studentů, ze třetího ročníku pouhých 6 studentů. Z programu EM opakuje informatiku 13 lidí a z programu SII opakují 2 studenti.

6.2.2.4 Studium v zahraničí

Cílová, vysvětlovaná proměnná je zde jasně stanovena - studium v zahraničí (x_{54}). Jako vstupní proměnné byly z matice S vybrány proměnné, které by mohly studium v zahraničí ovlivnit: typ studia (x_{32}), ročník studia (x_{33}), studijní program (x_{34}), přijetí ke studiu (x_{36}), typ střední školy (x_{37}), prospěchové stipendium (x_{38}), opakování předmětů (x_{39}), preferování předtermínů (x_{49}) a zkouškové termíny (x_{50}), využívání jazykového centra (x_{51}), studovaný jazyk (x_{52}), certifikát (x_{53}), náročnost studia na FES (x_{75}), vytyčené cíle (x_{76}), komunikační schopnosti (x_{78}), komunikační schopnosti v cizím jazyce (x_{79}), přístup k učení (x_{81}), návštěva psychologické poradny (x_{82}).

Results for output field X54

Individual Models

- Comparing \$C-X54 with X54

'Partition'	1_Training		2_Testing	
Correct	340	93,92%	130	97,74%
Wrong	22	6,08%	3	2,26%
Total	362		133	
- Comparing \$R-X54 with X54

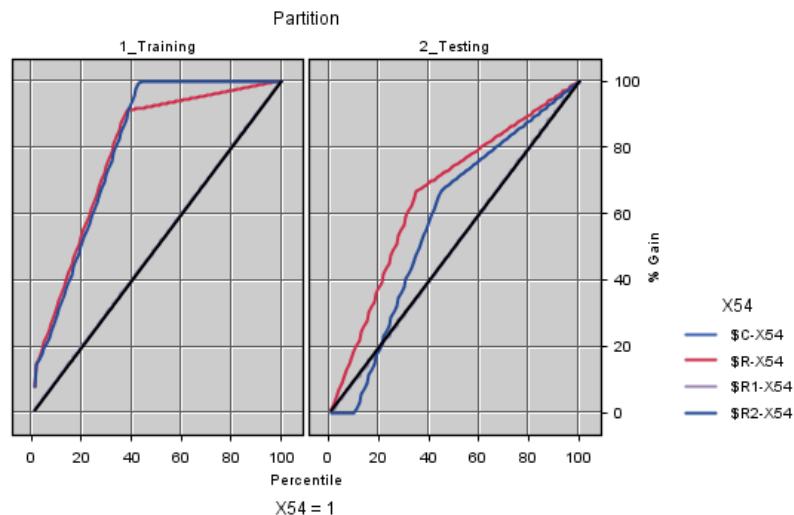
'Partition'	1_Training		2_Testing	
Correct	341	94,2%	130	97,74%
Wrong	21	5,8%	3	2,26%
Total	362		133	
- Comparing \$R1-X54 with X54

'Partition'	1_Training		2_Testing	
Correct	340	93,92%	130	97,74%
Wrong	22	6,08%	3	2,26%
Total	362		133	
- Comparing \$R2-X54 with X54

'Partition'	1_Training		2_Testing	
Correct	341	94,2%	130	97,74%
Wrong	21	5,8%	3	2,26%
Total	362		133	

Obrázek 38 - Vyhodnocení kvality výstupů RS (x_{54}) [vlastní]

Podle ohodnocení kvality výstupu uzlem Analysis (Obrázek 38) nelze usoudit, jaký z výstupů bude nejkvalitnější. V tomto případě má mnohem větší vypovídací schopnost graf Evaluation (Obrázek 39). Křivky dvou uzlů (C&RT a CHAID) mají optimálnější tvar, než křivky ostatní.



Obrázek 39 - Výstup uzlu Evaluation (x_{54}) [vlastní]

Pro klasifikaci je vybrán RS uzlu C&RT (Příloha č. 13). Z množiny testovacích dat je 340 studentů, kteří v průběhu studia nevyužili možnosti studia na partnerských zahraničních univerzitách. Naprostá menšina, 22 studentů se dále člení na 17 studentů navazujícího magisterského studia a 5 studentů bakalářského studia. Z pěti studentů, kteří vycestovali do zahraničí a studují bakalářský studijní program, jsou v prvním a druhém ročníku dva studenti a ve třetím ročníku našli odvahu vycestovat tři studenti. Přičemž oba studenti prvního a druhého ročníku o sobě mohou říci, že jejich komunikační znalosti v cizím jazyce jsou na dobré úrovni. Na druhou stranu, je zde 206 studentů bakalářského programu (1 a 2 ročník), kteří nevycestovali do zahraničí a tvrdí o sobě, že jejich komunikační schopnosti jsou průměrné či slabé.

6.2.2.5 Vytyčené cíle

Pro klasifikaci studentů je nejprve použit expertní odhad a následně je na data aplikován uzel Feature Selection, který vyhodnotil, jaké atributy jsou důležité vzhledem k výstupní proměnné. Na závěr jsou výsledky porovnány.

Expertní odhad

Vstupem do modelu podle expertního odhadu jsou určeny následující atributy: typ studia (x_{32}), ročník studia (x_{33}), studijní program (x_{34}), prospěchové stipendium (x_{38}), opakování předmětů (x_{39}), preferování předmětů (x_{49}), zkouškové termíny (x_{50}), využívání jazykového centra (x_{51}), studium v zahraničí (x_{54}), studentský email (x_{74}), náročnost studia na FES (x_{75}), platové představy (x_{77}), doporučení studia přátelům (x_{80}), přístup k učení (x_{81}),

návštěva psychologické poradny (x_{82}). Podle Obrázku 40 je jako nejkvalitnější model označen uzel C5.0 s nejvhodnějším poměrem testovacích dat.

Results for output field X76

Individual Models

Comparing \$C-X76 with X76

'Partition'	1_Training		2_Testing	
Correct	247	69,38%	79	59,4%
Wrong	109	30,62%	54	40,6%
Total	356		133	

Comparing \$R-X76 with X76

'Partition'	1_Training		2_Testing	
Correct	241	67,7%	74	55,64%
Wrong	115	32,3%	59	44,36%
Total	356		133	

Comparing \$R1-X76 with X76

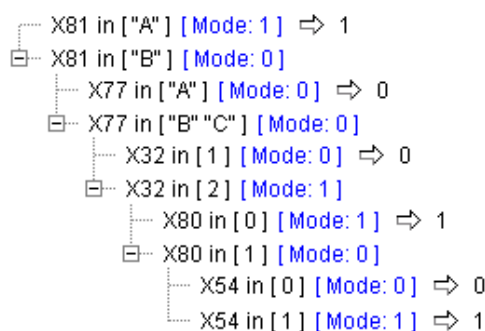
'Partition'	1_Training		2_Testing	
Correct	217	60,96%	73	54,89%
Wrong	139	39,04%	60	45,11%
Total	356		133	

Comparing \$R2-X76 with X76

'Partition'	1_Training		2_Testing	
Correct	223	62,64%	73	54,89%
Wrong	133	37,36%	60	45,11%
Total	356		133	

Obrázek 40 - Vyhodnocení kvality výstupů RS (x_{76}) [vlastní]

Výstup v podobě rozhodovacích pravidel u tohoto uzlu není už jednoznačně nejlogičtější. Podle expertního odhadu byl zvolen pro klasifikaci uzel C&RT, jehož rozhodovací pravidla jsou uvedena na Obrázku 41.



Obrázek 41 - Rozhodovací pravidla uzlu C&RT (x_{76}) [vlastní]

Klasifikace je realizována pomocí pěti atributů. Prvním atributem je přístup k učení, dále platové představy studenta, typ studia, doporučení studia přátelům a posledním atributem je studium v zahraničí.

Uzel Feature Selection

Jako nejdůležitější atributy podle uzlu Faecture Selection byly vyhodnoceny typ studia (x_{32}), typ střední školy (x_{37}), design portálu Stag (x_{68}), platové představy (x_{77}), komunikační schopnosti (x_{78}), komunikační schopnosti v cizím jazyce (x_{79}), přístup k učení (x_{81}).

Results for output field X76

Individual Models

Comparing \$C-X76 with X76

'Partition'	1_Training		2_Testing	
Correct	237	66,57%	72	54,14%
Wrong	119	33,43%	61	45,86%
Total	356		133	

Comparing \$R-X76 with X76

'Partition'	1_Training		2_Testing	
Correct	249	69,94%	74	55,64%
Wrong	107	30,06%	59	44,36%
Total	356		133	

Comparing \$R1-X76 with X76

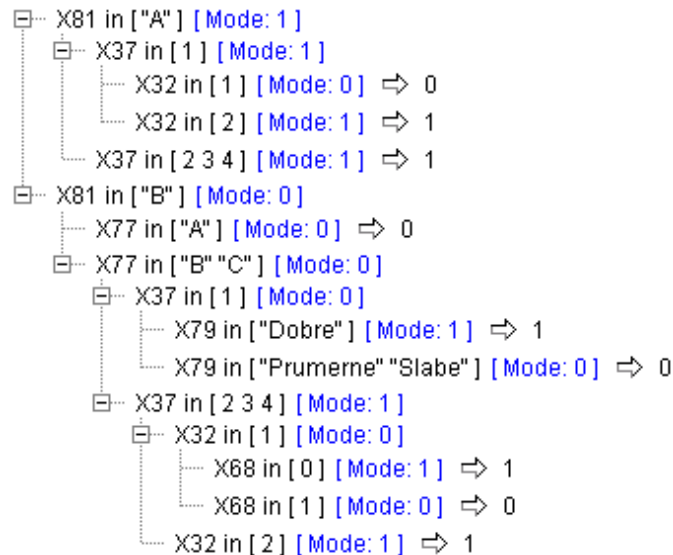
'Partition'	1_Training		2_Testing	
Correct	217	60,96%	73	54,89%
Wrong	139	39,04%	60	45,11%
Total	356		133	

Comparing \$R2-X76 with X76

'Partition'	1_Training		2_Testing	
Correct	244	68,54%	72	54,14%
Wrong	112	31,46%	61	45,86%
Total	356		133	

Obrázek 42 - Vyhodnocení kvality výstupů RS (x₇₆) [vlastní]

Jak je vidět na Obrázku 42 za nejkvalitnější výstup lze považovat výstup C&RT. Rozhodovací pravidla tohoto uzlu jsou uvedena na Obrázku 43.



Obrázek 43 - Rozhodovací pravidla uzlu C&RT [vlastní]

Větvení probíhá podle atributů přístup k učení, typ střední školy, typ studia, platové představy, komunikační schopnosti v cizím jazyce a spokojenost s designem portálu Studijní agendy (Stag).

Vyhodnocení klasifikace

Vybrané metody jsou porovnány pomocí chybných hodnot v trénovací a testovací množině dat (Tabulka 16). Vybrána je položka s minimálními hodnotami.

Tabulka 16 - Přesnost klasifikace [vlastní]

Použité metody	Počet chybných hodnot	
	Trénovací množina	Testovací množina
Expertní odhad	115	59
Feature Selection	107	59

Z výše uvedeného pozorování (Tabulka 16) vyplývá, že způsob klasifikace na základě vstupních atributů vyhodnocených uzlem Feature Selection je vhodnější. Výsledkem klasifikace je rozhodovací strom uvedený v Příloze č. 14. Tento RS klasifikuje studenty na 186 studentů, kteří mají stanovený cíl a 110 studentů bez stanoveného cíle. Prvním kritériem dělení je přístup k učení. Studentů, kteří mají stanovený cíl a učí se se zájmem o obor je 83. O něco více (103 studentů) se učí jen když musí. Ti co se učí se zájmem o obor a mají vytyčené cíle, jsou v šestnácti případech z gymnázií a v šedesátisedmi případech z ostatních středních škol. Z gymnazistů je 5 studentů na bakalářském typu studia a 11 na navazujícím magisterském. U studentů, kteří mají stanovené cíle a učí se, jen když musí, je klasifikace provedena podle jiných atributů. Těchto 103 studentů je rozděleno podle platových představ na skupinu s nízkými představami o platě (10-20 tisíc Kč) zastoupenou čtrnácti studenty a skupinu s vyššími platovými představami (20 tisíc Kč a více) zastoupenou 89 studenty. Z těchto 89 je 24 studentů z gymnázií a 65 z ostatních středních škol. Zarážejícím faktem je, že pouze 3 studenti z gymnázií s vyššími platovými představami a vytyčeným cílem do budoucna označili své schopnosti komunikace v cizím jazyce za dobré. Ostatních 21 studentů označilo své komunikační schopnosti v cizím jazyce za průměrné či slabé.

6.3 Shluková analýza

Shluková analýza je provedena na propojených datových maticích O a S. Po použití uzlu Kohonen a ponecháním primárního nastavení je vytvořeno 9 shluků. Vzhledem k počtu proměnných a jejich struktuře je obtížné hledat zákonitosti a jednoznačně popsat každý shluk. Po nastavení parametrů pro shlukovou analýzu v záložce Expert jsou vygenerovány čtyři shluky. V tomto menším počtu shluků je hledání závislostí a popsání shluků již mnohem logičtější, i když neustále velmi podrobné.

Model Kohonen také přiřadí důležitost jednotlivým atributům. Většina atributů je vyhodnocena jako důležitá. Najdou se také atributy určené jako středně důležité (pohlaví, účast na volitelné TV, vlastník živnostenského listu) a nedůležité (platba studia, brigáda, volný čas).

Nejpočetnější shluk zastoupený 246 studenty má souřadnice $X = 0$ a $Y = 0$. Druhý nejvíce zastoupený shluk (141 studentů) má souřadnice $X = 3$ a $Y = 0$. Méně početný je shluk $X = 1$ a $Y = 0$ s 92 studenty a na posledním místě je s 16 studenty shluk o souřadnicích $X = 2$ a $Y = 0$. Jednotlivé shluky jsou charakterizovány podle atributů, které byly vyhodnoceny shlukovou analýzou jako důležité.

Nejpočetnější shluk je tvořen téměř 50% dotázaných studentů. Z uvedených atributů vyplývá, že studenti zařazení do této skupiny v 51 % bydlí v Pardubicích. Bydlí buď na privátě, nebo v bytě v Pardubicích. Absolutní převahu mají studenti z Pardubického kraje. Přesto, že studenti nebydlí na kolejích, stravují se všichni v menze. S cenou menzy jsou studenti v 60% spokojeni. Fakt, že cena odpovídá kvalitě služeb je však popřen u 54 % studentů. S otvírací dobou menzy je spokojeno 80 % studentů. Velmi podobná hladina spokojenosti je u systému volného výběru jídel bez předchozího objednání, spokojeno je 85 % studentů. Katedru TV využívá 67 % studentů. Celkem uspokojivý je i fakt, že 65 % studentů je spokojeno s okolím univerzity. 76 % studentů se vyslovilo negativně k situaci s parkováním v okolí UPCE. Protože 71 % těchto studentů není vlastníkem automobilu, není možné přesně říci do jaké míry je tato nespokojenost vypovídající. Co se týká studijního programu, v 69% je shluk zastoupen studenty bakalářského typu. 47% tvoří studenti druhého ročníku, značná část je zastoupena také studenty prvního ročníku, třetí ročník je zastoupen minimálně (9 %). Věkový průměr je 21 let.

Druhý nejpočetněji zastoupený shluk (141 studentů) tvoří studenti, kteří opět nebydlí na kolejích. Ze 43 % jsou tito studenti z Pardubického kraje. Velké zastoupení (32 %) mají i studenti z Královéhradeckého kraje. Nejspíše tato situace ovlivňuje následující atribut - bydliště v Pardubicích. Studenti v Pardubicích nebydlí, ale denně dojíždí v 55%. Zbýlých 45% studentů bydlí na privatě, či bytě v Pardubicích. V menze se v tomto shluku nestravují žádní studenti. 53 % studentů využívá služeb katedry TV. Téměř stejná situace nastává v hodnocení spokojenosti s okolím UPCE a hodnocením parkování zde, jako v předchozím shluku. Spokojena s okolím UPCE je nadpoloviční většina studentů (57 %). S parkováním je nespokojeno 74 % studentů, přičemž 84 % studentů nevlastní automobil. Větší polovina studentů (52 %) je z bakalářského typu studia, kdy 45 % jsou studenti druhého ročníku, 30 % prvního ročníku a 24 % studentů třetího ročníku. Věkový průměr je zde o něco vyšší než v předchozím shluku - 22 let.

Třetím shlukem je skupina 92 studentů. 20 % studentů pochází z kraje Vysočina, dále 17 % studentů je z Královéhradeckého kraje, 15 % ze středočeského kraje a po 10% je zastoupení studentů z kraje Pardubického a Olomouckého. Tito studenti jsou již z 96% ubytováni na vysokoškolských kolejích. Co se týká spokojenosti na vysokoškolských kolejích, studenti nejsou spokojeni s cenou v 80%. Hodnocení toho, zda-li cena kolejí odpovídá poskytovaným službám je také celkem negativní. 87 % studentů vyslovilo nespokojenost s tímto faktem. Naopak se stávajícím systémem přidělování kolejí je 58 % studentů spokojeno. Všichni studenti z této skupiny se stravují v menze. Spokojenost studentů s menzou je podobná jako v prvním klastru. 57 % studentů je spokojena s cenou, že cena neodpovídá kvalitě jídla je přiřazeno 56% studentů. Téměř všichni studenti (93 %) jsou spokojeni s otvírací dobou menzy, obdobně i 90 % studentů je spokojeno se systémem volného výběru jídel. 72 % studentů v tomto klastru využívá služeb katedry TV. Studenti jsou v 55% spokojeni s parkováním v okolí UPCE a v 71% jsou spokojeni i s okolím UPCE. Opět vlastníky automobilu nejsou studenti z 87 %. Ohledně studijního programu jsou v tomto klastru zařazeni následující studenti: 70 % studenti bakalářského studijního typu, 50 % studenti prvního ročníku. Věkový průměr je 21 let.

Poslední shluk je zastoupen šestnácti studenty. Tito studenti jsou převážně z Libereckého kraje a Vysočiny, bydlí na vysokoškolských kolejích a v menze se nestravují. Jsou to převážně studenti druhého a třetího ročníku bakalářského studia s věkovým průměrem 22 let.

Závěr

Diplomová práce se zabývá problematikou klasifikace studentů FES. Cílem práce bylo provést tuto klasifikaci na základě realizovaného dotazníkového šetření, které bylo provedeno mezi studenty bakalářských a navazujících magisterských studijních programů.

Záměrem vytvořeného dotazníku bylo získat od studentů všeobecné informace, týkající se jednak jejich osobních záležitostí tak i jejich studijního programu. Získána byla od studentů i data ohledně hodnocení spokojenosti (studijních podmínek) na FES. Sběr probíhal na přednáškách a seminářích jednotlivých ročníků. Vzhledem k aplikované metodě sběru dat, byla návratnost dotazníků téměř stoprocentní – 95 %.

Po nasbírání potřebných informací od studentů byly vytvořeny datové matice. Matice O a S byly následně analyzovány v prostředí SPSS Clementine 10.1. Matice H je uvedena pouze v Příloze č. 7 této práce. Matice O a S byly zpracovány nejdříve samostatně. Po stanovení cílových proměnných (spokojenost s cenou kolejí, spokojenost s cenou jídel v menze, financování studentů, opakování předmětů – matematických a infromatických, studium v zahraničí a vytyčené cíle studentů) proběhlo modelování závislosti na výstupech. Ze všech vstupních atributů byly na základě expertního odhadu stanoveny jen ty důležité pro modelování. Následné modelování probíhalo čtyřmi algoritmy RS. Porovnáním výsledků byl vybrán vhodný RS, na jehož základě byly vyhodnoceny závislosti vstupních atributů na cílových proměnných.

V případech, kde expertní odhad podal příliš jednoduchý, případně nelogický výsledek byly pro výběr vstupních proměnných použity alternativní metody. Výsledky těchto alternativních metod se staly vstupními hodnotami pro RS. Výsledky RS, byly porovnány a opět stanoven výsledný rozhodovací strom.

Na závěr byla na propojené matice O a S aplikována shluková analýza. Výsledkem této analýzy byly čtyři shluky klasifikující studenty především podle bydliště v Pardubicích, hodnocení kolejí, stravování v menze s následným hodnocením a vyjádření studentů o spokojenosti s parkováním a okolím UPCE. Ve shlucích byly jako důležité atributy hodnoceny i typ studia, ročník a věk studentů.

Charakteristika těchto klastrů byla provedena pomocí atributů vyhodnocených touto metodou jako důležitých.

Zpracováním dotazníkového šetření za účelem klasifikace studentů vyšla najevo některá zajímavá fakta. Mnohé věci, které byly v této diplomové práci zkoumány lze ovlivnit jak ze strany studentů, tak ze strany vedení FES.

Seznam literatury a použitých zdrojů

Literatura

- [1] BERKA, Petr. *Dobývání znalostí z databází*. 1 vyd. Praha: Academia, 2003. 366 s. ISBN 80-200-1062-9.
- [2] HAN, Jiawei, KAMBER, Micheline. *Data Mining, Concepts and Techniques*. 770 s. 2. vyd. San Francisco: Morgan Kautmann, 2006. ISBN 1-55860-901-6.
- [3] KOZEL, Roman. *Moderní marketingový výzkum: nové trendy, kvantitativní a kvalitativní metody a techniky, průběh a organizace, aplikace v praxi, přínosy a možnosti*. 1 vyd. Praha: Grada, 2006. 277 s. ISBN 80-247-0966-X.
- [4] KUBANOVÁ, Jana. *Statistické metody pro ekonomickou a technickou praxi*. Bratislava: Statis, 2003. 247 s. ISBN 80-85659-31-X
- [5] MELOUN, Milan, MILITKÝ, Jiří. *Kompendium statistického zpracování*. 1 vyd. Praha: Academia, 2002. 764 s. ISBN 80-200-1008-4.
- [6] PETR, Pavel. *Data mining*. 1 vyd. Pardubice: Univerzita Pardubice, 2006. 144 s. ISBN 80-7194-886-1.
- [7] ROKACH, L.; MAIMON, O. *Data mining with decision trees – Theory and applications*. Singapore: World Scientific Printes, 2008. 244. ISBN 981-277-171-9.
- [8] RUD, Olivia Parr, MAQERA, Ivo, DANĚK, Milan. *Data mining, praktický průvodce dolováním dat pro efektivní prodej, cílený marketing a podporu zákazníků (CRM)*. 1 vyd. Praha: Compuerr Press, 2001. 329 s. ISBN 80-7226-577-6.
- [9] ŘEZANKOVÁ, Hana. *Analýza dat z dotazníkových šetření*. 1 vyd. Praha: Professional Publishing, 2007. 212 s. ISBN 978-80-86946-49-8.
- [10] ŘEZANKOVÁ, Hana, HÚSEK, Dušan, SNÁŠEL, Václav. *Shluková analýza*. 2. rozš. vyd. Praha: Professional Publishing, 2009. 218 s. ISBN 978-80-86946-81-8.
- [11] *SPSS Inc. Clementine (r) 7.0 User's Guide*. 2002. 741 s. ISBN 1-56827-295-2 .
- [12] ZBOŘIL, Kamil. *Marketingový výzkum*. 1 vyd. Praha: Vysoká škola ekonomická, 1994. ISBN 80-7079-389-9.

Ostatní zdroje

- [13] *Department of cybertic* [online]. [cit. 2009-03-30]. Český. Dostupný z WWW: <<http://cyber.felk.cvut.cz/gerstner/teaching/zbd/DataMining1-hout.pdf>>
- [14] *Dotazník* [online]. [cit. 2009-03-30]. Český. Dostupný z WWW: <<http://www.dotaznik-online.cz>>
- [15] *Univerzita Pardubice: FES:Historie* [online]. [cit. 2009-03-30]. Český. Dostupný z WWW: <www.upce.cz/fes/fakulta/o-fakulte/historie.html>
- [16] *ČSÚ* [online]. [cit. 2009-03-30]. Český. Dostupný z WWW: <http://www.czso.cz/csu/klasifik.nsf/i/co_je_statisticka_klasifikace>
- [17] *Výroční zpráva 2007* [online]. [cit. 2009-03-30]. Český. Dostupný z WWW: <<http://www.upce.cz/fes/deska/zakladni-dokumenty/vyrocní-zprava.html>>

Seznam tabulek

Tabulka 1 – Ukázka datového slovníku Osobni_dotazovani [vlastní]	24
Tabulka 2 - Ukázka datového slovníku Studijni_program [vlastní].....	25
Tabulka 3 – Zastoupení studentů v krajích [vlastní]	27
Tabulka 4 – Počet studentů jednotlivých oborů [vlastní]	29
Tabulka 5 – Oblíbené kluby v Pardubicích [vlastní].....	29
Tabulka 6 – Ubytování studentů [vlastní]	30
Tabulka 7 – Spokojenost s univerzitními kolejemi [vlastní].....	31
Tabulka 8 – Spokojenost studentů s menzou [vlastní]	32
Tabulka 9 - Počet studentů opakujících předmět [vlastní]	33
Tabulka 10 – Počet opakovaných předmětů [vlastní]	33
Tabulka 11 – Destinace navštěvované studenty [vlastní].....	34
Tabulka 12 - Využívání Univerzitní knihovny [vlastní]	35
Tabulka 13 – Účel návštěvy Univerzitní knihovny [vlastní].....	35
Tabulka 14 – Přesnost klasifikace [vlastní].....	56
Tabulka 15 – Konečné porovnání metod [vlastní]	56
Tabulka 16 - Přesnost klasifikace [vlastní]	68

Seznam obrázků

Obrázek 1 – Organizační struktura FES [17]	12
Obrázek 2 - Vývoj počtu studentů [vlastní].....	13
Obrázek 3 – Vývoj počtu studentů kombinovaného a prezenčního studia [vlastní]	13
Obrázek 4 – Typy dotazování [3]	16
Obrázek 5 – Úspěšnost sběru dat [vlastní]	22
Obrázek 6 – Zastoupení pohlaví [vlastní]	27
Obrázek 7 – Počty studentů prezenčního a kombinovaného studia [vlastní]	28
Obrázek 8 – Počty studentů jednotlivých ročníků [vlastní]	28
Obrázek 9 – Spokojenost s vysokoškolskými kolejemi [vlastní]	31
Obrázek 10 – Spokojenost s menzou [vlastní]	32
Obrázek 11 – Počet studentů opakující jednotlivé oblasti [vlastní]	33
Obrázek 12 – Studium v zahraničí [vlastní]	34
Obrázek 13 – Frekvence využívání Univerzitní knihovny [vlastní].....	35
Obrázek 14 – Začlenění modelování do dobývání znalostí, upraveno podle[1]	36
Obrázek 15 – Popis modelu [vlastní]	37
Obrázek 16 – Schéma neuronové sítě [10].....	45
Obrázek 17 – Uzel Var. File, karta Types po načtení dat [vlastní]	46
Obrázek 18 – Uzel Data Audit ze souboru Studijni_program [vlastní]	47
Obrázek 19 – Rozhodovací pravidla uzlu QUEST (x_6) [vlastní]	49
Obrázek 20 – Rozhodovací pravidla uzlu QUEST (x_6) [vlastní]	50
Obrázek 21 - Rozhodovací pravidla uzlu QUEST(x_6) [vlastní].....	50
Obrázek 22 - Rozhodovací pravidla uzlu QUEST(x_6) [vlastní].....	51
Obrázek 23 – Vyhodnocení kvality výstupů RS (x_{11}) [vlastní].....	52
Obrázek 24 - Rozhodovací pravidla uzlu CHAID (x_{11}) [vlastní].....	52
Obrázek 25 - Vyhodnocení kvality výstupů RS (x_{11}) [vlastní]	53
Obrázek 26 - Rozhodovací pravidla uzlu CHAID (x_{11}) [vlastní].....	53

Obrázek 27 - Vyhodnocení kvality výstupů RS (x_{11}) [vlastní]	54
Obrázek 28 - Rozhodovací pravidla uzlu C5.0 (x_{11}) [vlastní].....	54
Obrázek 29 - Vyhodnocení kvality výstupů RS (x_{11}) [vlastní]	55
Obrázek 30 - Rozhodovací pravidla uzlu C&RT (x_{11}) [vlastní].....	55
Obrázek 31 - Vyhodnocení kvality výstupů RS (Financování) [vlastní]	58
Obrázek 32 – Vyhodnocení kvality výstupů RS (x_{39}) [vlastní].....	59
Obrázek 33 – Výstup uzlu Evaluation (x_{39}) [vlastní]	60
Obrázek 34 – Vyhodnocení kvality výstupů RS (x_{41}) [vlastní].....	61
Obrázek 35 – Výstup uzlu Evaluation (x_{41}) [vlastní]	62
Obrázek 36 – Vyhodnocení kvality výstupů RS (x_{43}) [vlastní].....	63
Obrázek 37 – Výstup uzlu Evaluation (x_{43}) [vlastní]	63
Obrázek 38 - Vyhodnocení kvality výstupů RS (x_{54}) [vlastní]	64
Obrázek 39 - Výstup uzlu Evaluation (x_{54}) [vlastní].....	65
Obrázek 40 - Vyhodnocení kvality výstupů RS (x_{76}) [vlastní]	66
Obrázek 41 - Rozhodovací pravidla uzlu C&RT (x_{76}) [vlastní].....	66
Obrázek 42 - Vyhodnocení kvality výstupů RS (x_{76}) [vlastní]	67
Obrázek 43 - Rozhodovací pravidla uzlu C&RT [vlastní].....	67

Seznam zkratek

Pojem	Význam
ECS	Ekonomika a celní správa
EK	Ekonomika pro kriminalisty
EM	Ekonomika a management
FES	Fakulta ekonomicko-správní
HPS	Hospodářská politika a správa
IBS	Informační a bezpečnostní systémy
RS	Rozhodovací stromy
RR	Regionální rozvoj
SII	Systémové inženýrství a informatika
SOM	Samoorganizující se mapy
STAG	Studijní agenda
TV	Tělesná výchova
UK	Univerzitní knihovna
UPCE	Univerzita Pardubice
VES	Veřejná ekonomika a správa

Seznam příloh

Příloha č. 1 - Studijní programy (obory) prezenčního a kombinovaného studia

Příloha č. 2 - Ukázka dotazníku - Část A

Příloha č. 3 - Ukázka dotazníku – Část B

Příloha č. 4 - Ukázka dotazníku – Část C

Příloha č. 5 - Datový slovník – matice O

Příloha č. 6 - Datový slovník – matice S

Příloha č. 7 - Ukázka datové matice Hodnoceni_spokojenosti

Příloha č. 8 - Spokojenost s cenou jídla v menze – rozhodovací strom

Příloha č. 9 - Financování studia – rozhodovací strom

Příloha č. 10 - Opakování předmětů – rozhodovací strom

Příloha č. 11 - Opakování matematických předmětů - rozhodovací strom

Příloha č. 12 - Opakování předmětů informatiky – rozhodovací strom

Příloha č. 13 - Studium v zahraničí – rozhodovací strom

Příloha č. 14 - Vytyčené cíle studentů – rozhodovací strom

Přílohy

Příloha č. 5 - Studijní programy (obory) prezenčního a kombinovaného studia

Kód stud. progr.	Typ studijního programu	Název studijního programu	Název studijního oboru	Forma studia	Standardní doba studia v ak. letech
B6209	bakalářský	Systémové inženýrství a informatika	Informatika ve veřejné správě	prezenční kombinovaná	3,5
B6209	bakalářský	Systémové inženýrství a informatika	Informatika ve veřejné správě	prezenční kombinovaná	3
B6209	bakalářský	Systémové inženýrství a informatika	Regionální a informační management	prezenční, kombinovaná	3
B6209	bakalářský	Systémové inženýrství a informatika	Regionální a informační management	prezenční AJ	3
B6209	bakalářský	Systémové inženýrství a informatika	Informační a bezpečnostní systémy	prezenční, kombinovaná	3
M6209	magisterský	Systémové inženýrství a informatika	Informatika ve veřejné správě	prezenční	5
N6209	navazující magisterský	Systémové inženýrství a informatika	Informatika ve veřejné správě	prezenční, kombinovaná	2
N6209	navazující magisterský	Systémové inženýrství a informatika	Regionální a informační management	prezenční, kombinovaná	2
N6209	navazující magisterský	Systémové inženýrství a informatika	Regionální a informační management	prezenční AJ	2
N6209	navazující magisterský	Systémové inženýrství a informatika	Pojistné inženýrství	prezenční	2
P6209	doktorský	Systémové inženýrství a informatika	Informatika ve veřejné správě	prezenční, kombinovaná	3

Kód stud. progr.	Typ studijního programu	Název studijního programu	Název studijního oboru	Forma studia	Standardní doba studia v ak. letech
P6209	doktorský	System Engineering and Informatics	Informatics within Public Administration	prezenční, kombinovaná AJ	3
B6202	bakalářský	Hospodářská politika a správa	Veřejná ekonomika a správa	prezenční, kombinovaná	3
B6202	bakalářský	Hospodářská politika a správa	Ekonomika a celní správa	kombinovaná	3
B6202	bakalářský	Hospodářská politika a správa	Ekonomika pro kriminalisty	kombinovaná	3
N6202	navazující magisterský	Hospodářská politika a správa	Ekonomika veřejného sektoru	prezenční, kombinovaná	2
N6202	navazující magisterský	Hospodářská politika a správa	Regionální rozvoj	prezenční	2
P6202	doktorský	Hospodářská politika a správa	Regionální a veřejná ekonomie	prezenční, kombinovaná	3
P6202	doktorský	Economic Policy and Administration	Regional and Public Economics	prezenční, kombinovaná AJ	3
B6208	bakalářský	Ekonomika a management	Management podniku	prezenční, kombinovaná	3
B6208	bakalářský	Ekonomika a management	Ekonomika a provoz podniku	prezenční, kombinovaná	3
B6208	bakalářský	Ekonomika a management	Management ochrany podniku a společnosti	prezenční, kombinovaná	3
B6208	navazující magisterský	Ekonomika a management	Ekonomika a management podniku	prezenční, kombinovaná	2
B6208	doktorský	Ekonomika a management	Management	prezenční, kombinovaná	3

Příloha č. 6 - Ukázka dotazníku - Část A

Část A - Osobní dotazování (identifikace studenta)

1. Jaké je Vaše pohlaví?

muž žena

2. Kolik je Vám let?

19 20 21 22 23 24 a více

3. Jaký je Váš rodinný stav?

svobodný /á ženatý/vdaná rozvedený/rozvedená

4. Z jakého Kraje pocházíte?

Karlovarský kraj	<input type="checkbox"/>	Středočeský kraj	<input type="checkbox"/>	Jihomoravský kraj	<input type="checkbox"/>
Plzeňský kraj	<input type="checkbox"/>	Praha	<input type="checkbox"/>	Olomoucký kraj	<input type="checkbox"/>
Ústecký kraj	<input type="checkbox"/>	Královéhradecký kraj	<input type="checkbox"/>	Moravskoslezský kraj	<input type="checkbox"/>
Jihočeský kraj	<input type="checkbox"/>	Pardubický kraj	<input type="checkbox"/>	Zlínský kraj	<input type="checkbox"/>
Liberecký kraj	<input type="checkbox"/>	Vysočina	<input type="checkbox"/>		

5. Kde bydlíte v Pardubicích?

Vysokoškolské koleje Nebydlím v Pardubicích, denně dojíždím
Privát, byt v Pardubicích

Pokud bydlíte na kolejích, odpovězte prosím, na následující otázky:

- | | | | |
|---|---|-----|----|
| a | Jste spokojeni s cenou univerzitních kolejí | ANO | NE |
| b | Odpovídá cena kolejí službám, které jsou poskytovány | ANO | NE |
| c | Vyhovuje Vám systém přidělování kolejí | ANO | NE |
| d | Souhlasíte se systémem propojení účtu kolejí a účtu menzy | ANO | NE |

6. Využíváte služeb stravování v menze?

ANO NE

Pokud jste odpověděli ano, odpovězte prosím, na následující otázky:

- | | | | |
|---|---|-----|----|
| a | Jste spokojeni s cenou jídel | ANO | NE |
| b | Odpovídá cena jídel jejich kvalitě | ANO | NE |
| c | Jste spokojeni s otvírací dobou menzy | ANO | NE |
| d | Vyhovuje Vám systém volného výběru jídel bez objednání | ANO | NE |
| e | Využíváte nabídky doplňkového sortimentu (studená kuchyně, saláty, dezerty,...) | ANO | NE |

7*. Využíváte nabídky katedry TV a sportu?

ANO Povinná TV (v rámci výuky) NE, nevyužívám
Volitelná TV (v rámci výuky)
Výcviková forma (mimo výuky)

8*. Kdo platí Vaše náklady za studium?

Rodiče Sám/a Stipendium

9. Chodíte během školního roku (říjen - červen) na brigádu?

ANO NE

Pokud jste odpověděli ANO, kolik hodin týdně pracujete?

1 až 5 5 až 10 10 až 20 20 a více

10. Jste vlastníkem živnostenského listu?

ANO NE

11. Máte k dispozici vlastní automobil?

ANO NE

Část A - Osobní dotazování (identifikace studenta)

12. Je dle Vašeho názoru situace s parkováním v okolí Univerzity vyhovující?

ANO NE

13. Jste spokojen/a s okolím kolejí a Univerzity?

ANO NE

14* Jak trávíte volný čas?

sport četba kulturní akce

15. Jaké jsou nejoblíbenější kluby v Pardubicích, které navštěvujete?

.....

Příloha č. 7 - Ukázka dotazníku – Část B

Část B - Studijní program

16. Typ studia
Denní Dálkové
17. Studijní program
Bakalářský Navazující magisterský Doktorský
18. Ročník studia
1 2 3 4 5
19. Studijní program
Hospodářská politika a správa
Ekonomika a management
Systémové inženýrství a informatika
20. Studijní obor
- | | | |
|-----|--|--------------------------|
| HPS | Veřejná ekonomika a správa | <input type="checkbox"/> |
| | Ekonomika pro kriminalisty | <input type="checkbox"/> |
| | Ekonomika a celní správa | <input type="checkbox"/> |
| | Ekonomika veřejného sektoru | <input type="checkbox"/> |
| | Regionální rozvoj | <input type="checkbox"/> |
| | Regionální a veřejná ekonomie | <input type="checkbox"/> |
| EAM | Management podniku | <input type="checkbox"/> |
| | Management ochrany podniku a společnosti | <input type="checkbox"/> |
| | Ekonomika a provoz podniku | <input type="checkbox"/> |
| | Ekonomika a management podniku | <input type="checkbox"/> |
| SII | Informatika ve veřejné správě | <input type="checkbox"/> |
| | Regionální a informační management | <input type="checkbox"/> |
| | Informační a bezpečnostní systémy | <input type="checkbox"/> |
| | Pojistné inženýrství | <input type="checkbox"/> |
21. Ke studiu na FES jste byl/a přijat/a:
bez přijímacího řízení na základě přijímacího řízení na odvolání
22. Jaká je Vaše střední škola?
Gymnazium Obchodní akademie Průmyslová škola
Jiná střední škola
23. Jste příjemcem prospěchového stipendia?
ANO NE
24. Opakujete v tomto školním roce předměty z minulých semestrů?
ANO 1 předmět NE
2 předměty
3 a více předmětů
- 25*. Předměty které opakujete, se týkají následujících oblastí:
matematika ekonomie informatika
- 26*. Preferované oblasti studia (které předměty Vás zajímají) jsou:
matematika ekonomie informatika
27. Jste spokojen/a s nabídkou volitelných (alternativních) předmětů?
ANO NE
28. Jste spokojen/a se systémem tvorby rozvrhu?
ANO NE
29. Vítáte možnosti splnění zkoušek v předetermínech?
ANO NE
30. Snažíte se ukončit ročník (LS) v řádných zkuškových termínech nebo spíše odkládáte zkoušky na období srpen - září?
řádné termíny srpen - září

Část B - Studijní program

31. Využíváte nabídky jazykového centra?

ANO NE

Pokud jste odpověděl/a ano, odpovězte prosím, na následující otázky:

a* Jaký jazyk studujete AJ NJ FJ RJ Jiné
b Složil/a jste některý z mezinárodně uznávaných ANO NE
certifikátů nabízených JC

32. Využil/a jste možnosti studia v zahraničí na některé z partnerských univerzit?

ANO NE

Pokud jste odpověděl/a ano, odpovězte prosím, na následující otázky:

a V rámci jakého programu jste vycestoval/a DAAD Erasmus Jiné
b V jaké zemi jste studoval/a
c V jakém jazyce jste studoval/a
d Měl tento pobyt pro Vás přínos ANO NE
e Jak dlouhý byl pobyt na zahraniční univerzitě ZS LS ZS+LS 3 a více semestrů

33. Využíváte služeb univerzitní knihovny?

ANO NE

Pokud jste odpověděl/a ano, odpovězte prosím, na následující otázky:

a* Za jakým účelem navštěvuješ knihovnu studium internet, PC denní tisk, časopisy
b Jak často chodíš do knihovny 1x týdně vícekrát týdně 1x měsíčně
c Vyhovuje Vám otvírací doba knihovny ANO NE
d Vyhovuje Vám vybavení knihovny studijní literaturou ANO NE

34. Jste spokojen/a s novým designem informačního systému STAG?

ANO NE

a Vyhovoval Vám lépe původní design ANO NE
b Máte návrh na změnu IS Stag

35. Jste spokojen/a s novým designem webových stránek Upce?

ANO NE

a Vyhovoval Vám lépe původní design ANO NE
b Máte návrh na změnu webových těchto stránek ANO NE

36. Sledujete průběžně aktuální informace o studiu (termíny odevzdání seminárních prací, směrnice k BP či DP,...)

ANO NE

37. Používáte studentský email?

ANO NE

38. Studium na FES shledáváte:

Náročné středně těžké lehké

39. Máte vytyčené cíle - profesní zaměření (v jakém oboru chcete po studiu pracovat)?

ANO NE

40. Jaké jsou Vaše platové představy po dokončení studia?

10 - 20 tis. 20 - 30 tis. 30 tis. a více

41. Vaše komunikační schopnosti jsou dle Vašeho názoru:

dobré průměrné slabé

42. Vaše komunikační schopnosti v cizím jazyce jsou dle Vašho názoru:

dobré průměrné slabé

43. Doporučil /a byste studium na FES UPCE svým přátelům?

ANO NE

44. Učíte se:

ze zájmu o obor jen když musím

45. Navštívil/a jste psychologickou poradnu na naší škole v případě potřeby?

ANO NE

Příloha č. 8 - Ukázka dotazníku – Část C

Část C - Hodnocení spokojenosti

Ohodnoťte spokojenost s následujícími kritérii stupnicí od 1 do 5 :

1 = velmi spokojen, 2 = spíše spokojen, 3 = spokojen, 4 = spíše nespokojen, 5 = nespokojen

46. Náplň oboru	Spokojenost					
a povědomí o cílech oboru	1	2	3	4	5	
b obsah oboru (skladba předmětů)	1	2	3	4	5	
d možnost pracovního uplatnění	1	2	3	4	5	
e dostupnost zaměstnání	1	2	3	4	5	
f rozsah splnění vašich představ o oboru	1	2	3	4	5	
g rozsah přípravy na zaměstnání	1	2	3	4	5	
47. Pedagogický sbor						
a spolehlivost učitelů (přesnost, dodržení termínů)	1	2	3	4	5	
b schopnost vyučovat	1	2	3	4	5	
c vědomostní základ a zručnost	1	2	3	4	5	
d vhodnost použitých vyučovacích a studijních technik	1	2	3	4	5	
e ochota učitelů pomoci	1	2	3	4	5	
f dohled nad Vaší prací	1	2	3	4	5	
g možnost styku s vyučujícím mimo časy přednášek /cvičení	1	2	3	4	5	
48. Ohodnoťte:						
a informace o místě a konání ZK	1	2	3	4	5	
b flexibilita termínů v odevzdávání prací	1	2	3	4	5	
c informace o kriteriích hodnocení	1	2	3	4	5	
d důslednost známkovacích měřítek	1	2	3	4	5	
e použité metody hodnocení	1	2	3	4	5	
f rychlost vrácení hodnocenných výsledků	1	2	3	4	5	
g užitečnost zpětné diskuse o úkolech	1	2	3	4	5	
49. Zdroje pro výuku						
a zajištění studijních materiálů	1	2	3	4	5	
b dostupné vybavení pro studium a výuku	1	2	3	4	5	
c stav učeben	1	2	3	4	5	
d velikost učeben a počet studentů ve skupinách	1	2	3	4	5	
e speciální místnosti (pro skupinové výuky, prezentace...)	1	2	3	4	5	
50. Informace poskytované studijním oddělením						
Před zahájením Vašeho studia na Univerzitě						
a den otevřených dveří	1	2	3	4	5	
b informace obsažené v letácích	1	2	3	4	5	
c informace o oboru než jste nastoupil/la na Univerzitu (předměty)	1	2	3	4	5	
d přijímací řízení (realizace, organizace,..)	1	2	3	4	5	
e informace o studiu	1	2	3	4	5	
Po zahájení Vašeho studia na Univerzitě						
a příručky pro studenty	1	2	3	4	5	
c proces podání požadavků	1	2	3	4	5	
d úřední hodiny studijního oddělení	1	2	3	4	5	
e projednávání vašich požadavků	1	2	3	4	5	

Část C - Hodnocení spokojenosti

51. Počítačové zařízení - knihovna

a	možnost použití počítačů, kdykoli chcete	1	2	3	4	5
b	jak staré jsou počítače	1	2	3	4	5
c	množství softwaru, jenž je dostupný	1	2	3	4	5
d	pomoc od personálu v těchto prostorách	1	2	3	4	5
e	spolehivost počítačové sítě	1	2	3	4	5
f	rychlost internetu	1	2	3	4	5
g	možnost tisku	1	2	3	4	5
h	cena tisku a kopírování	1	2	3	4	5

52. Počítačové zařízení - tiskové centrum

a	možnost použití počítačů, kdykoli chcete	1	2	3	4	5
b	jak staré jsou počítače	1	2	3	4	5
c	množství softwaru, jenž je dostupný	1	2	3	4	5
d	pomoc od personálu v těchto prostorách	1	2	3	4	5
e	spolehivost počítačové sítě	1	2	3	4	5
f	rychlost internetu	1	2	3	4	5
g	možnost tisku	1	2	3	4	5
h	cena tisku a kopírování	1	2	3	4	5

Příloha č. 5 – Datový slovník – matice O

Název atributu	Datový typ	Rozsah hodnot	Kód - Hodnota atributu		Stručný popis
x1	Flag	Z/M	M	Muž	pohlaví
			Z	Žena	
x2	Range	[19 - 24]	19	19	věk
			20	20	
			21	21	
			22	22	
			23	23	
			24	24 a více	
x3	Set	[1,2,3]	1	Svobodný/á	rodinný stav
			2	Ženatý/vdaná	
			3	Rozvedený/á	
x4	Set	[1 - 15]	1	Karlovarský kraj	kraj
			2	Plzeňský kraj	
			3	Ústecký kraj	
			4	Jihočeský kraj	
			5	Liberecký kraj	
			6	Středočeský kraj	
			7	Praha	
			8	Královhradecký kraj	
			9	Pardubický kraj	
			10	Vysočina	
			11	Jihomoravský kraj	
			12	Olomoucký kraj	
			13	Moravskoslezský kraj	
			14	Zlínský kraj	
			15	Slovensko	
x5	Set	[0,1,2,3,4]	1	Koleje	bydliště v Pardubicích
			2	Privát, doma	
			3	Dojíždí	
x6	Set	[0,1,2]	0	Ne	spokojenost s cenou kolejí
			1	Ano	
			2	Neodpověděl	
x7	Set	[0,1,2]	0	Ne	cena odpovídá službám kolejí
			1	Ano	
			2	Neodpověděl	
x8	Set	[0,1,2]	0	Ne	spokojenost s přidělováním kolejí
			1	Ano	
			2	Neodpověděl	
x9	Set	[0,1,2]	0	Ne	souhlas s propojením účtu kolejí a menzy
			1	Ano	
			2	Neodpověděl	
x10	Flag	0/1	0	Ne	stravování v menze
			1	Ano	
x11	Set	[0,1,2]	0	Ne	spokojenost s cenou jídel
			1	Ano	
			2	Neodpověděl	
x12	Set	[0,1,2]	0	Ne	cena jídel = kvalita jídel
			1	Ano	
			2	Neodpověděl	

Název atributu	Datový typ	Rozsah hodnot	Kód - Hodnota atributu		Stručný popis
x13	Set	[0,1,2]	0	Ne	spokojenost s otvírací dobou
			1	Ano	
			2	Neodpověděl	
x14	Set	[0,1,2]	0	Ne	spokojenost se systémem volného výběru jídel
			1	Ano	
			2	Neodpověděl	
x15	Set	[0,1,2]	0	Ne	využívá-li student doplňkový sortiment menzy
			1	Ano	
			2	Neodpověděl	
x16	Flag	0/1	0	Ne	využívá-li student katedru TV
			1	Ano	
x17	Set	[0,1]	0	Ne	povinná TV
			1	Ano	
x18	Set	[0,1]	0	Ne	volitelná TV
			1	Ano	
x19	Set	[0,1]	0	Ne	výcviková forma
			1	Ano	
x20	Set	[A,AB,ABC,AC,B,BC,C,D]	A	Rodiče	platba studia
			B	Sám/a	
			C	Stipendium	
			D	Neodpověděl	
x21	Flag	0/1	0	Ne	brigáda
			1	Ano	
x22	Set	[A,B,C,D,E]	A	1 až 5	brigáda (hodin /týden)
			B	5 až 10	
			C	10 až 20	
			D	20 a více	
			E	Neodpověděl	
x23	Flag	0/1	0	Ne	vlastník ŽL
			1	Ano	
x24	Flag	0/1	0	Ne	vlastník automobilu
			1	Ano	
x25	Flag	0/1	0	Ne	vyhovující parkování v okolí UPCE
			1	Ano	
x26	Flag	0/1	0	Ne	spokojenost s okolím UPCE
			1	Ano	
x27	Flag	1/0	0	Ne	volný čas - sport
			1	Ano	
x28	Flag	1/0	0	Ne	volný čas - četba
			1	Ano	
x29	Flag	1/0	0	Ne	volný čas - kulturní akce
			1	Ano	
x30	Set	[310 - Zralok]			oblíbené kluby v Pardubicích

Příloha č. 6 – Datový slovník – matice S

Název atributu	Datový typ	Rozsah hodnot	Kód - Hodnota atributu		Stručný popis
x31	Flag	Prezencni/Kombinovane	Prezencni	Prezenční studium	forma studia
			Kombinovane	Kombinované studium	
x32	Flag	1/2	1	Bakalářský	typ studia
			2	Navazující	
X33	Ordered Set	[1,2,3]	1	První	ročník studia
			2	Druhý	
			3	Třetí	
			4	Čtvrtý	
			5	Pátý	
x34	Set	[EM, HPS,SII]	HPS	Hospodářská politika a správa	studijní program
			EM	Ekonomika a management	
			SII	Systémové inženýrství a informatika	
x35	Set	[EMP, ..., VES]	VES	Veřejná ekonomika a správa	studijní obor
			EK	Ekonomika pro kriminalisty	
			ECS	Ekonomika celní správa	
			EVS	Ekonomika veřejného sektoru	
			RR	Regionální rozvoj	
			RVE	Regionální a veřejná ekonomie	
			MP	Management podniku	
			MOPS	Management ochrany podniku a společnosti	
			EPP	Ekonomika a provoz podniku	
			EMP	Ekonomika a management podniku	
			IVS	Informatika ve veřejné správě	
			RIM	Regionální a informační management	
			IBS	Informační a bezpečnostní systémy	
x36	Set	[A,B,C]	A	Bez přijímacího řízení	přijetí
			B	Na základě přijímacího řízení	
			C	Na odvolání	
x37	Set	[1,2,3,4]	1	Gymnazium	typ střední školy
			2	Obchodní akademie	
			3	Průmyslová škola	
			4	Jiná škola	
x38	Flag	0/1	0	Ne	prospěchové stipendium
			1	Ano	
x39	Flag	0/1	0	Ne	opakování předmětů
			1	Ano	
x40	Set	[0,1,2,3]	0	neodpověděl	počet opakovaných předmětů
			1	Jeden	
			2	Dva	
			3	Tři	
x41	Flag	0/1	0	Ne	předměty v oblasti matematiky
			1	Ano	
x42	Flag	0/1	0	Ne	předměty v oblasti ekonomie
			1	Ano	
x43	Flag	0/1	0	Ne	předměty v oblasti informatiky
			1	Ano	
x44	Flag	0/1	0	Ne	preferovaná oblast matematika
			1	Ano	
x45	Flag	0/1	0	Ne	preferovaná oblast ekonomie
			1	Ano	
x46	Flag	0/1	0	Ne	preferovaná oblast informatika
			1	Ano	
x47	Flag	0/1	0	Ne	spokojenost s volitelnými
			1	Ano	
x48	Flag	0/1	0	Ne	spokojenost s tvorbou rozvrhu
			1	Ano	
x49	Flag	1/0	0	Ne	preferování předtermínů
			1	Ano	
x50	Flag	A/B	A	Řádné termíny (např. červen)	zkouškové termíny
			B	Pozdější termíny (např. září)	
x51	Flag	0/1	0	Ne	využívání jazykového centra
			1	Ano	

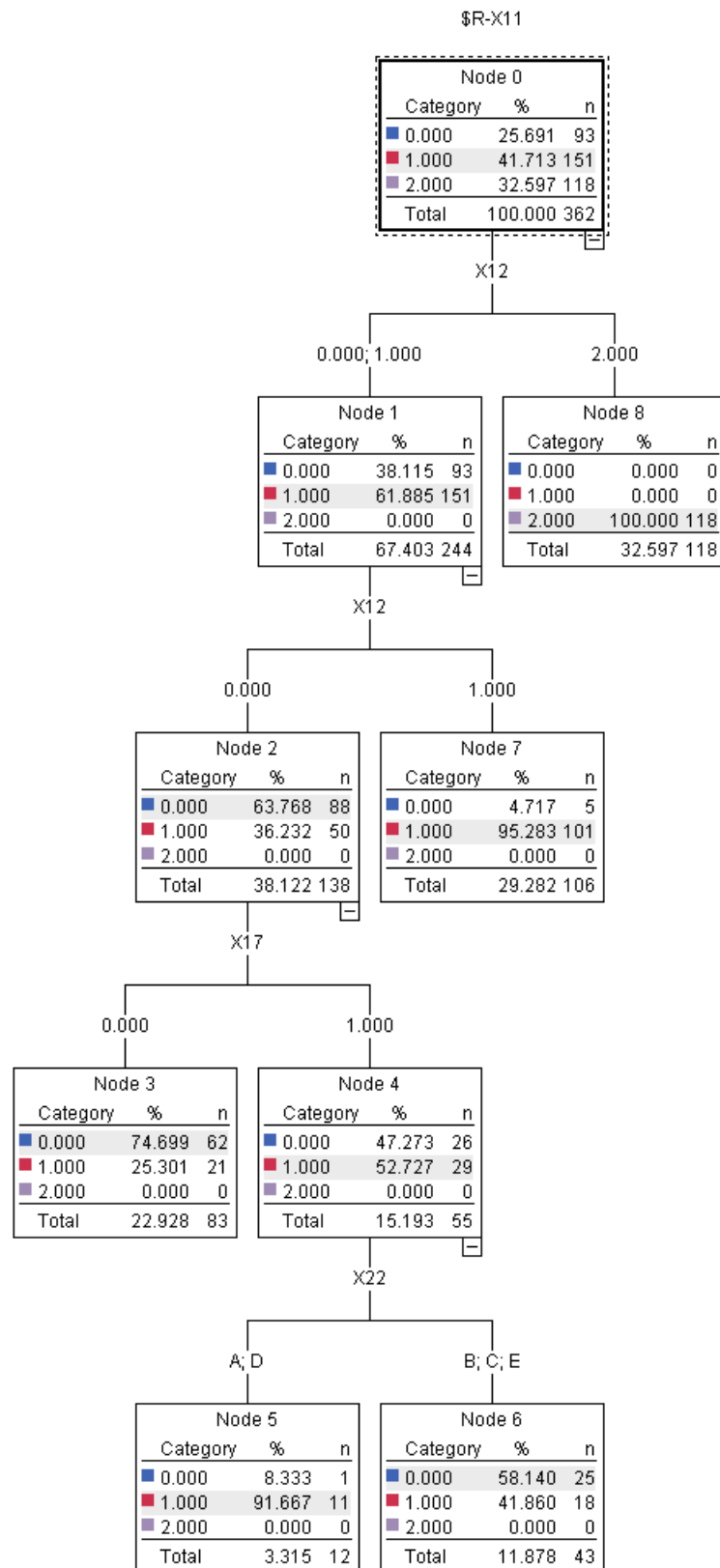
Název atributu	Datový typ	Rozsah hodnot	Kód - Hodnota atributu		Stručný popis
x52	Set	[AJ,NJ, ..., zadna]	AJ	Anglický jazyk	studovaný jazyk
			NJ	Německý jazyk	
			FJ	Francouzština	
			RJ	Ruský jazyk	
			Jiné	Jiné jazyky	
			Zadny	Žádný jazyk	
x53	Set	[0,1,2]	0	Ne	certifikát
			1	Ano	
			2	neodpověděl	
x54	Flag	0/1	0	Ne	studium v zahraničí
			1	Ano	
x55	Set	[0,2,3]	0	neodpověděl	progra vycestování
			1	DAAD	
			2	Erasmus	
			3	Jiné	
x56	Set	[Finsko, ..., zadna]	Finsko	Finsko	země studia
			Itálie	Itálie	
			Madeira	Madeira	
			Německo	Německo	
			Portugalsko	Portugalsko	
			Slovensko	Slovensko	
			Slovinsko	Slovinsko	
			Turecko	Turecko	
			zadna	neodpověděl	
x57	Set	[AJ,NJ, ..., zadna]	AJ	Anglický jazyk	jazyk studia
			NJ	Německý jazyk	
			SJ	Slovenský jazyk	
			PJ	Portugalština	
			zadna	neodpověděl	
x58	Set	[0,1,2]	0	Ne	přínos pobytu
			1	Ano	
			2	neodpověděl	
x59	Set	[LS, ..., zadna]	LS	Letní semestr	délka pobytu
			ZS	Zimní semestr	
			LS+ZS	Oba semestry	
			zadna	neodpověděl	
x60	Flag	0/1	0	Ne	využití univerzitní knihovny
			1	Ano	
x61	Set	[0,1,2]	0	Ne	knihovna - studium
			1	Ano	
			2	neodpověděl	
x62	Set	[0,1,2]	0	Ne	knihovna - inetrnet,PC
			1	Ano	
			2	neodpověděl	
x63	Set	[0,1,2]	0	Ne	knihovna - denní tisk, časopisy
			1	Ano	
			2	neodpověděl	
x64	Set	[A,B,C,D]	A	1x týdně	frekvence využívání knihovny
			B	Vícekrát týdně	
			C	1x měsíčně	
			D	neodpověděl	
x65	Set	[0,1,2]	0	Ne	spokojenost s otvárací dobou
			1	Ano	
			2	neodpověděl	
x66	Set	[0,1,2]	0	Ne	spokojenost s literaturou
			1	Ano	
			2	neodpověděl	
x67	Flag	0/1	0	Ne	spokojenost s designem portálu
			1	Ano	
x68	Flag	0/1	0	Ne	původní design lepší
			1	Ano	
x69	Set			návrhy na změnu	návrhy na změnu
x70	Flag	0/1	0	Ne	spokojenost s designem
			1	Ano	

Název atributu	Datový typ	Rozsah hodnot	Kód - Hodnota atributu		Stručný popis
x71	Flag	0/1	0	Ne	původní design lepší
			1	Ano	
x72	Set	[0,1,2]	0	Ne	návrhy na změnu
			1	Ano	
			2	neodpověděl	
x73	Flag	0/1	0	Ne	sledovanost aktuálních informací
			1	Ano	
x74	Flag	0/1	0	Ne	studentský email
			1	Ano	
x75	Ordered Set	[1,2,3]	1	Náročné	náročnost studia na FES
			2	Středně těžké	
			3	Lehké	
x76	Flag	0/1	0	Ne	vytyčené cíle
			1	Ano	
x77	Set	[A,B,C,D]	A	10 - 20 tis.	platové představy
			B	20 - 30 tis.	
			C	30 tis. a více	
			D	neodpověděl	
x78	Ordered Set	[Dobře,Prumerne,Slabe]	Dobře	Dobré	komunikační schopnosti
			Prumerne	Průměrné	
			Slabe	Slabé	
x79	Ordered Set	[Dobře,Prumerne,Slabe]	Dobře	Dobré	komunikační schopnosti v cizím jazyce
			Prumerne	Průměrné	
			Slabe	Slabé	
x80	Flag	0/1	0	Ne	doporučení studia přátelům
			1	Ano	
x81	Flag	A/B	A	Ze zájmu o obor	přístup k učení
			B	Jen když musím	
x82	Flag	0/1	0	Ne	návštěva psychologické
			1	Ano	

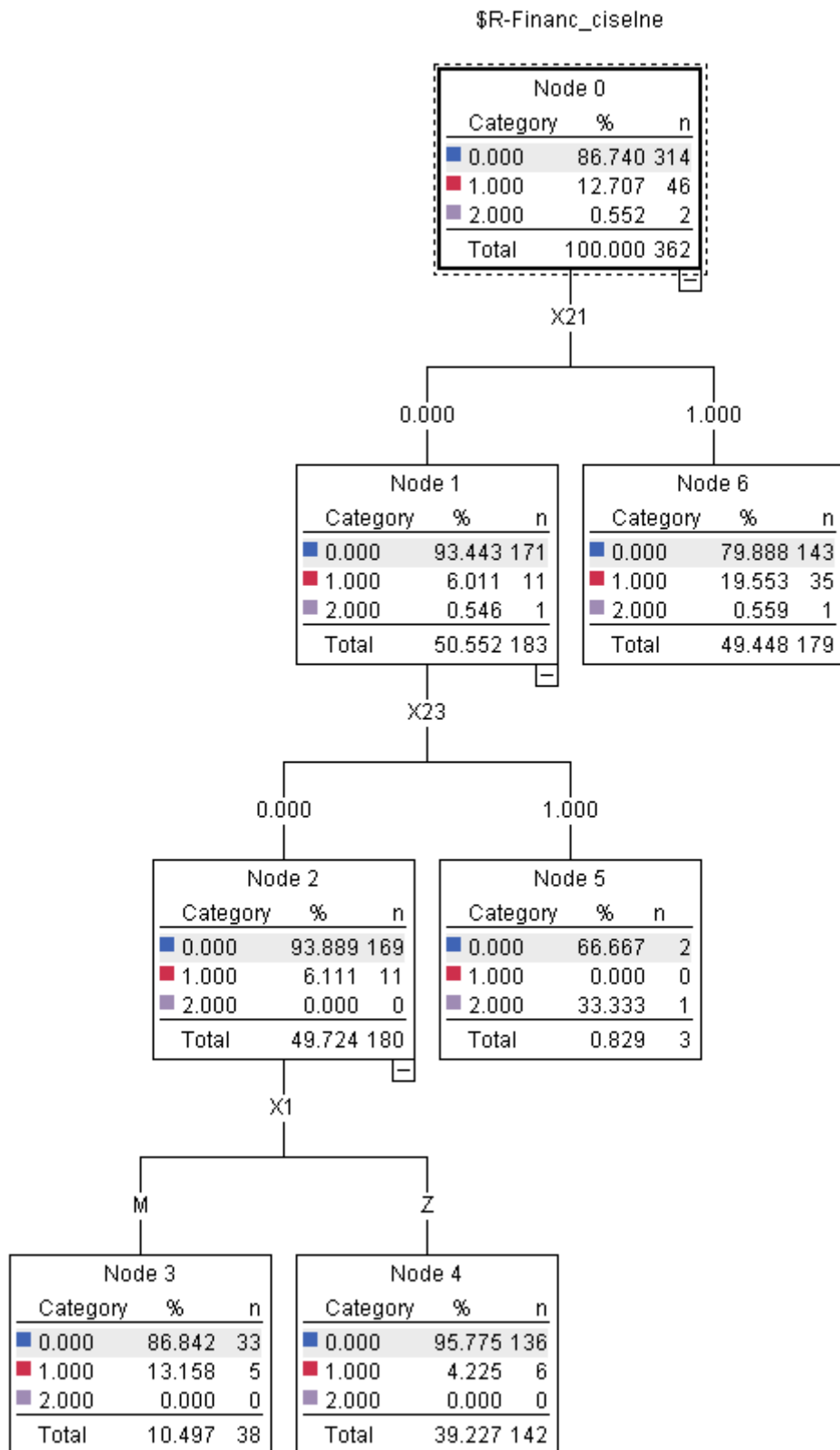
Příloha č. 7 – Ukázka datové matice Hodnoceni_spokojenosti

ID	S1	S2	S3	S4	S5	S6	S7	S8
1	3	3	3	3	4	4	3	3
2	2	4	2	3	2	3	1	1
3	2	4	2	2	3	3	4	5
4	3	4	4	4	4	4	3	3
5	3	4	3	4	3	3	4	3
6	1	3	3	3	4	4	2	3
7	3	3	3	3	3	3	2	4
8	2	3	2	2	2	2	1	3
9	5	4	4	4	4	4	1	2
10	5	4	3	3	5	4	1	2
11	2	3	2	3	3	2	3	4
12	3	3	3	5	5	3	4	2
13	3	4	4	3	4	3	1	2
14	4	5	1	2	4	4	2	2
15	2	2	3	3	3	3	3	3
16	3	3	4	4	4	4	2	2
17	3	3	3	3	4	3	2	3
18	2	2	2	3	4	2	3	2
19	3	3	2	2	3	2	2	3
20	2	2	2	3	2	2	2	3
21	1	2	2	2	4	2	5	4
22	2	3	2	2	2	3	2	2
23	3	3	3	3	3	3	2	4
24	4	4	2	2	3	3	3	3
25	3	3	2	2	3	2	3	3
26	1	4	1	1	3	2	1	2
27	3	3	2	1	2	2	2	3
28	2	4	1	1	2	3	2	2
29	3	3	2	1	3	2	2	2
30	2	3	2	4	3	4	2	3
31	3	3	2	2	2	2	2	1
32	1	2	2	2	2	1	2	2
33	2	2	3	3	2	4	3	4
34	2	2	3	3	3	4	3	4
35	4	4	5	4	5	5	4	5
36	3	4	2	4	3	3	3	4
37	3	2	4	3	3	2	3	4
38	3	3	3	3	3	3	3	3
39	3	3	3	3	5	5	3	3
40	2	2	1	1	1	1	2	3
41	2	2	3	2	1	2	2	2
42	2	3	3	2	2	4	2	3
43	4	4	3	3	4	4	4	4
44	2	1	1	3	4	4	4	3
45	3	4	3	3	3	3	3	3
46	3	4	2	3	3	4	4	3
47	2	1	2	2	2	1	2	2
48	2	3	2	3	3	4	2	3
49	3	3	2	4	3	4	4	3

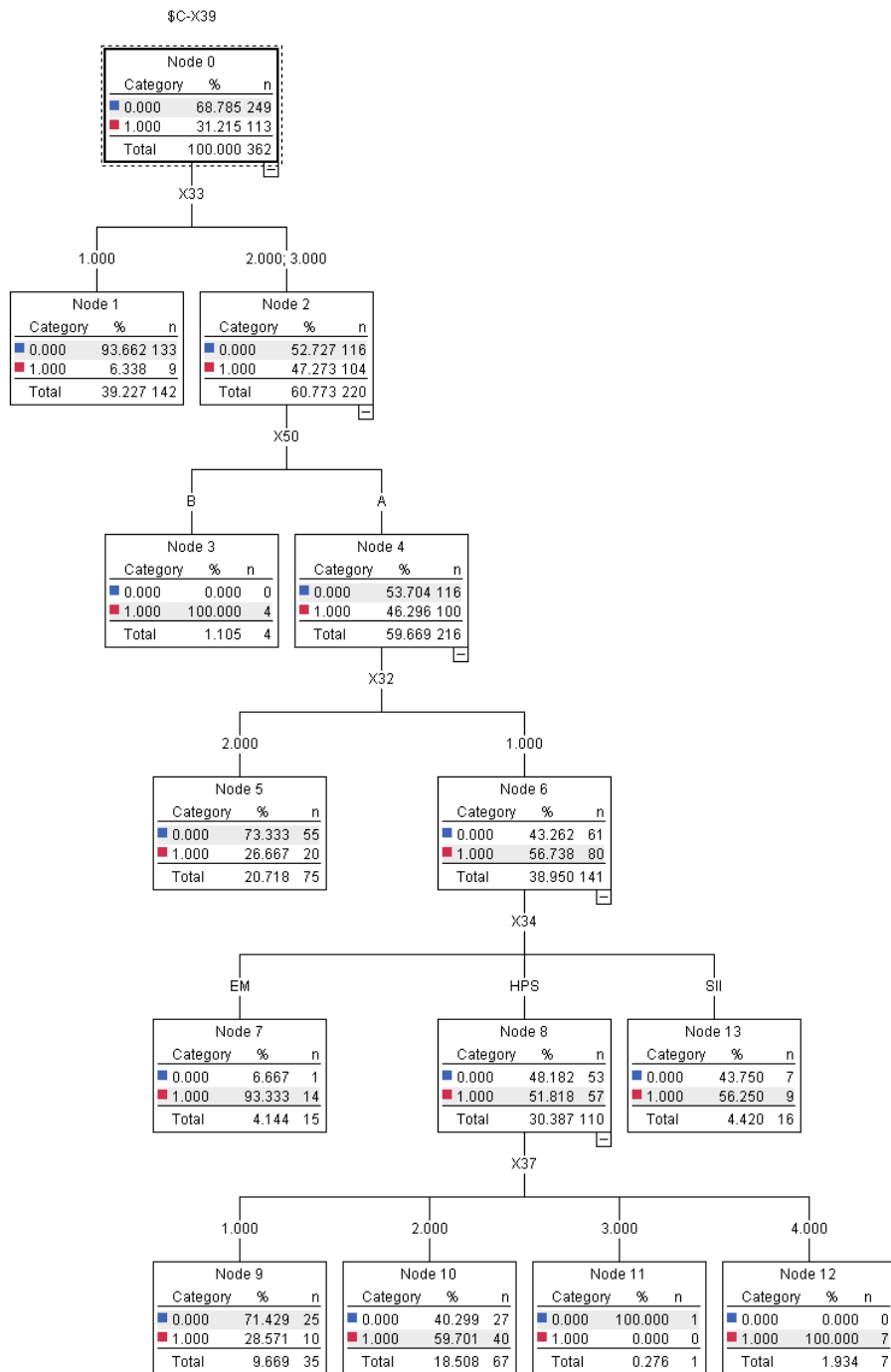
Příloha č. 8 - Spokojenost s cenou jídla v menze – rozhodovací strom



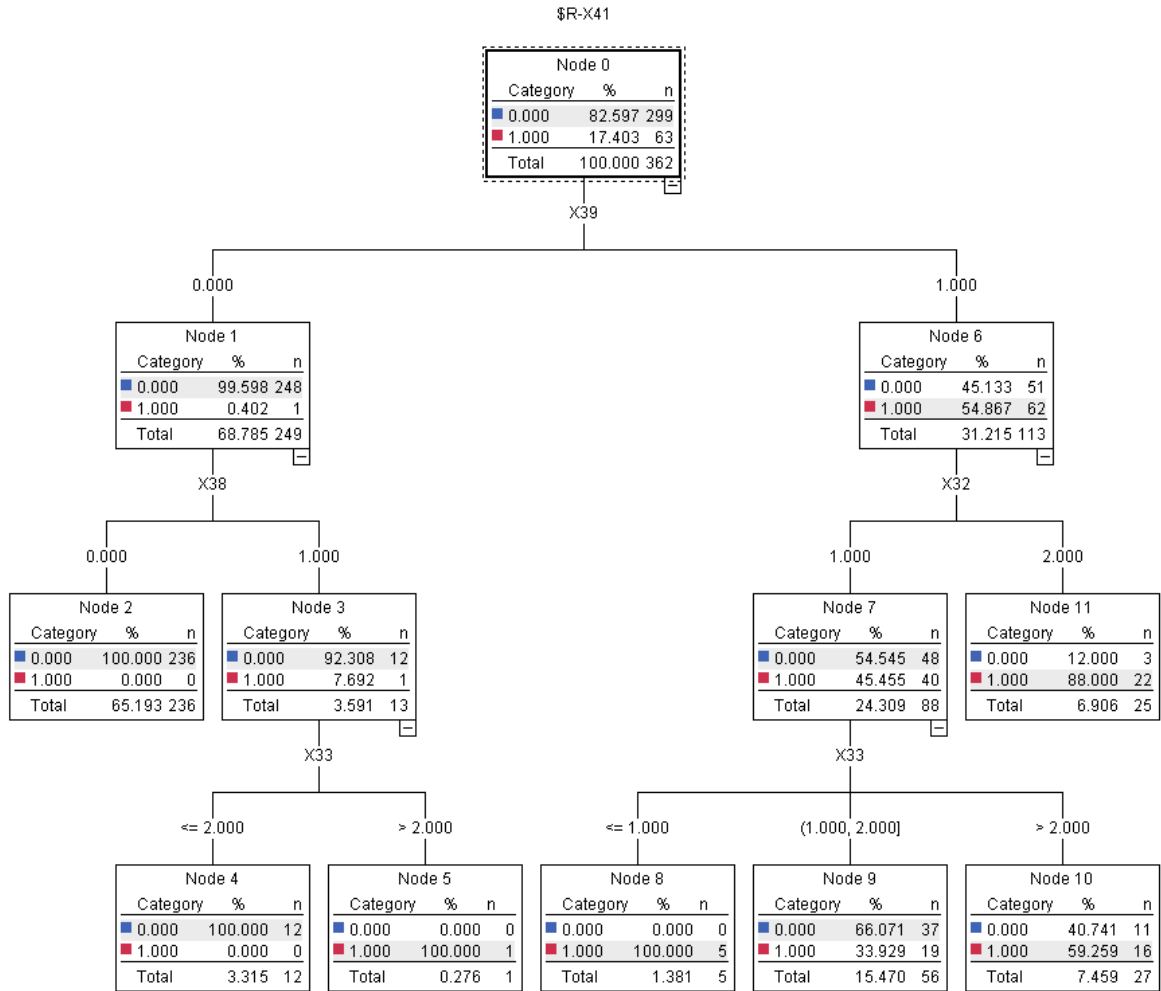
Příloha č. 9 – Financování studia – rozhodovací strom



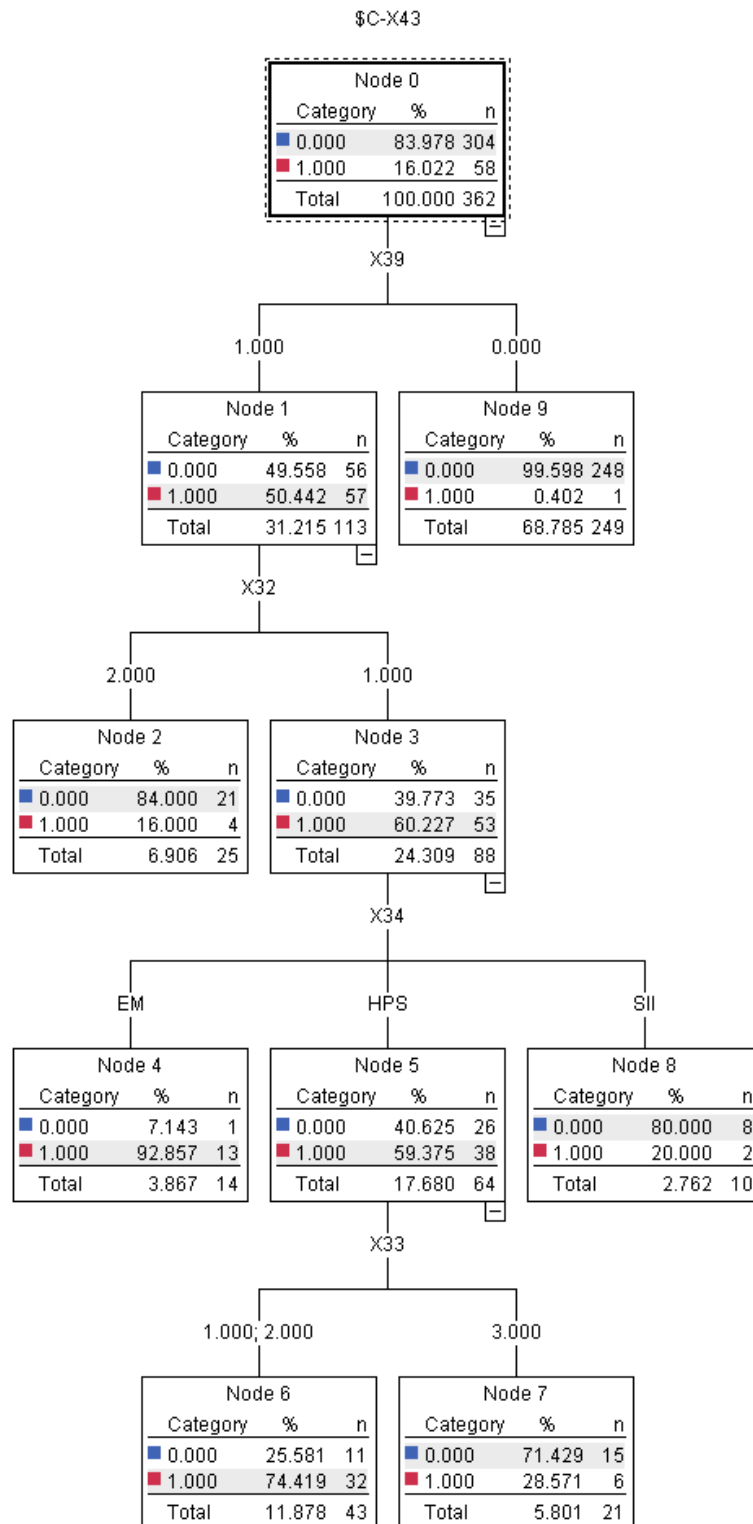
Příloha č. 10 – Opakování předmětů – rozhodovací strom



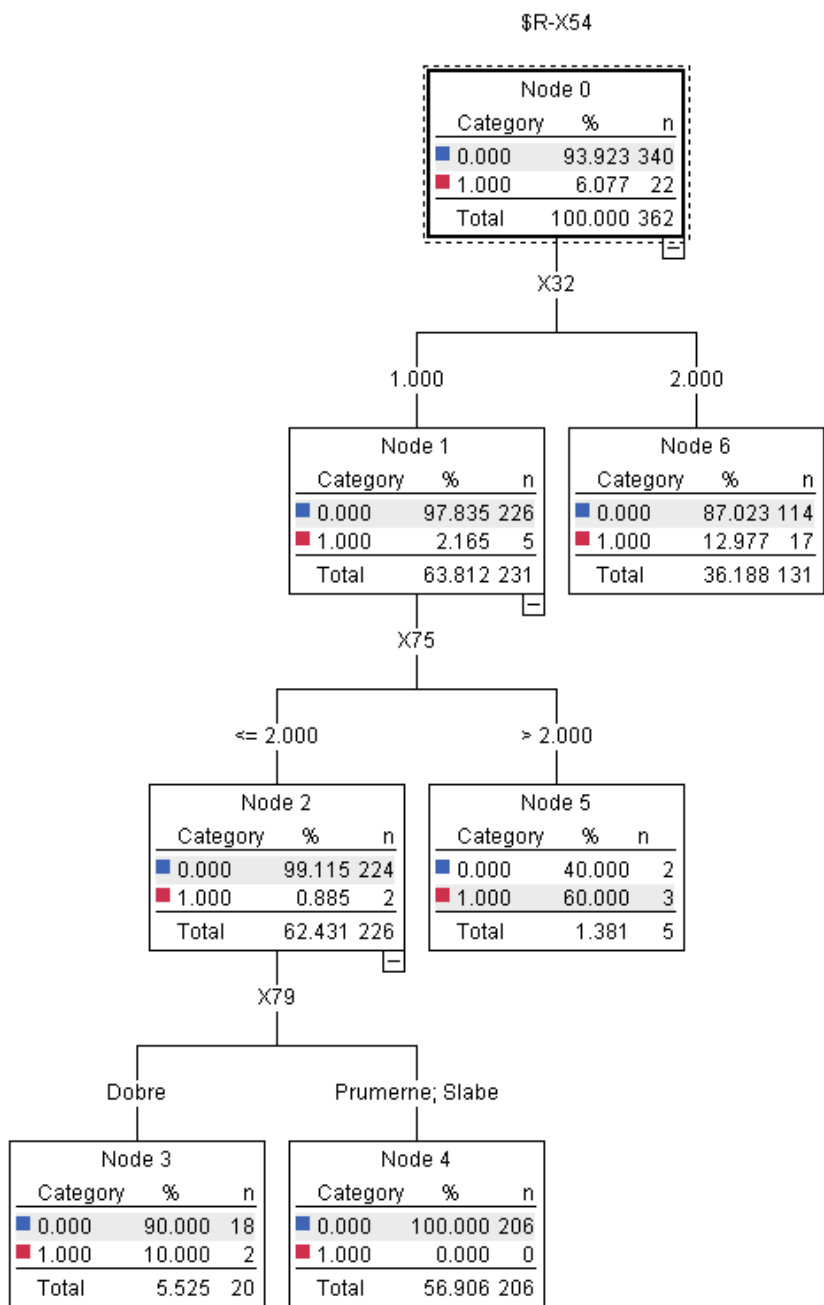
Příloha č. 11 – Opakování matematických předmětů - rozhodovací strom



Příloha č. 12 – Opakování předmětů informatiky – rozhodovací strom



Příloha č. 13 – Studium v zahraničí – rozhodovací strom



Příloha č. 14 – Vytčené cíle studentů – rozhodovací strom

