

**KLASTER ANALIZA USPJEHA STUDENATA NA FAKULTETU
INFORMACIJSKIH TEHNOLOGIJA U MOSTARU**

**CLUSTER ANALYSIS OF STUDENTS SUCCESS AT THE FACULTY
OF INFORMATION TECHNOLOGY IN MOSTAR**

dipl. ing. IT., Admir Šehidić, student²⁶⁴
dipl. ing. IT., Larisa Dedović, student²⁶⁵

Sažetak: *Na Fakultetu informacijskih tehnologija (FIT) u Mostaru, pored redovne nastave, pruža se mogućnost praćenja nastave po sistemu „Distance learning“. Distance learning podrazumijeva online distribuciju znanja uz primjenu savremenih informacijskih tehnologija. Nastavni plan i program po kojem studenti slušaju nastavu, pored opštih predmeta, zasnovan je na predmetima iz dvije aktuelne IT oblasti: softverski inženjering i sistemska i mrežna administracija. Osnovni cilj rada jeste, primjenom data mining tehnike klasteriranja, analizirati uspjeh studenata po oblastima, te utvrditi sličnosti i razlike u uspjehu redovnih i DL studenata. Korišteni set podataka zasnovan je na podacima prikupljenim dugogodišnjom upotrebom DL (Distance Learning Web Management System) sistema u kojem su pohranjeni podaci o studentima i njihovom uspjehu. Za potrebe istraživanja i prikaz rezultata korišten je Microsoft SQL Server Data Tools (SQL Server Analysis Service), alat namijenjen za napredne multidimenzionalne analize i data mining.*

Ključne riječi: *analiza uspjeha studenata, data mining, klasteriranje*

Abstract: *The Faculty of Information Technology (FIT) in Mostar, in addition to regular classes, provides the opportunity to study on a "Distance learning". Distance learning provides online distribution of knowledge by using modern information technologies. The curriculum, in addition to general subjects, based on the subjects of two most popular IT areas: software engineering and system and network administration. The primary goal of this work is, by applying clustering data mining techniques, analyze the success of students and identify similarities and differences in the success of regular and DL students. The used data set is based on the data collected by using DL platform (Distance Learning Web Management System) that stores data about students and their success. For the purposes of research and results presentation was used Microsoft SQL Server Data Tools (SQL Server Analysis Service), a tool designed for advanced multi-dimensional analysis and data mining.*

Key words: *student success analysis, data mining, clustering*

²⁶⁴ Fakultet informacijskih tehnologija Univerziteta „Džemal Bijedić“ u Mostaru, Sjeverni logor 12

²⁶⁵ Fakultet informacijskih tehnologija Univerziteta „Džemal Bijedić“ u Mostaru, Sjeverni logor 12

1. UVOD

Današnje informacijske sisteme karakterišu velike količine podataka nastale svakodnevnim poslovanjem. Često se govori o tome kako se kompanije „utapaju“ u podacima, a sa druge strane „gladuju“ za informacijama i znanjem. Za dobivanje korisnih informacija iz podataka se već odavno koristi statistika, dok je u zadnje vrijeme sve popularnija data mining disciplina koja uz pomoć računara omogućava pronalaženje i predviđanje korisnih informacija iz velikih skupova podataka.

U nastavku rada će, primjenom data mining tehnike klasteriranja, biti analiziran uspjeh studenata na Fakultetu informacijskih tehnologija u Mostaru. Prvenstveno, misli se na usporedbu studenata koji nastavu prate redovno u prostorijama Fakulteta u odnosu na studente koji studiraju po sistemu „distance learning“. Poseban osvrt u analizi će biti stavljen na predmete iz dvije ključne oblasti koje se izučavaju na Fakultetu: Softverski inženjering i Sistemska i mrežna administracija.

Osnovni ciljevi rada su: uočiti odstupanja u uspjehu između redovnih i DL studenata, ustanoviti da li postoje razlike u uspjehu po oblastima (nezavisno i zavisno od načina studiranja), te uvidjeti procenat kvalitetnih studenata tj. pronaći potencijalne stručnjake u oblasti (visoki prosjek u obje oblasti).

Rezultatima analize treba ustanoviti prednosti i nedostatke u načinu izvođenja nastave s ciljem unapređenja cjelokupnog nastavnog procesa, što će sigurno dovesti do većeg broja uspješnih studenata, a time i stručnjaka u pojedinim IT oblastima.

2. METODE

Data mining možemo definisati kao proces otkrivanja korisnih informacija iz velikih skupova podataka. Dosta je nova disciplina i do prije nekoliko godina primjenjivala se isključivo u naučnim krugovima, dok se u zadnje vrijeme sve više primjenjuje u različitim oblastima poslovanja kao što su marketing i trgovina, bankarstvo, medicina, obrazovanje i mnoge druge.



Admir Šehidić

Zvanje: Diplomirani inženjer informacijskih tehnologija

Tema diplomskog rada: „Unapređenje sistema za podršku učenju na daljinu – Administracijski modul“

Datum odbrane: 30.09.2011. god.

2012 – danas: Master studij Fakulteta informacijskih tehnologija u Mostaru, Smjer: Softverski inženjering

Projekti: QA Info Sys - Član tima za razvoj informacionog sistema za podršku internom osiguranju kvaliteta nastave na Univerzitetu „Džemal Bijedić“ u Mostaru, DLWMS v3 - Član tima za unapređenje i razvoj informacionog sistema za podršku učenju na daljinu, DLWMS v4 - Član tima za unapređenje i razvoj sistema za podršku učenju na daljinu (Distance Learning Web Management System version 4, UnSIS v5 - Član tima za unapređenje i razvoj informacionog sistema Univerziteta "Džemal Bijedić" u Mostaru (verzija 5)

Priznanja i nagrade: Rektorova nagrada Univerziteta "Džemal Bijedić" u Mostaru za postignuti uspjeh u akademskoj 2010/2011 godini.

Ova disciplina podrazumijeva korištenje različitih statističkih, matematičkih i programskih metoda za analizu podataka, gdje se postojeći (historijski) podaci obrađuju s ciljem otkrivanja skrivenih zakonitosti među njima, te predviđanje poslovanja. U tom procesu koriste se različite metode i tehnike koje se mogu svrstati u četiri osnovne skupine:

- Klasifikacijske metode,
- Asocijacijske metode,
- Klasteriranje (grupisanje),
- Numeričko predviđanje.

Za postizanje ciljeva istraživanja korištena je klaster analiza (klasteriranje).

Klaster analiza je naziv za grupu multivarijacionih tehnika čija je primarna svrha grupisanje objekata bazirano na karakteristikama koje posjeduju. Termin „klaster“ podrazumijeva skupinu objekata sličnih karakteristika, tako da je svaki objekat veoma sličan drugim objektima u klasteru, uz poštovanje unaprijed definisanog kriterijuma selekcije. Nastale skupine objekata bi trebale da pokažu visoku internu homogenost (unutar klastera), te visoku eksternu heterogenost (između klastera). Formalno, klastering struktura se može predstaviti kao skup podskupova $C = C_1, \dots, C_k$ od S , takvih da:

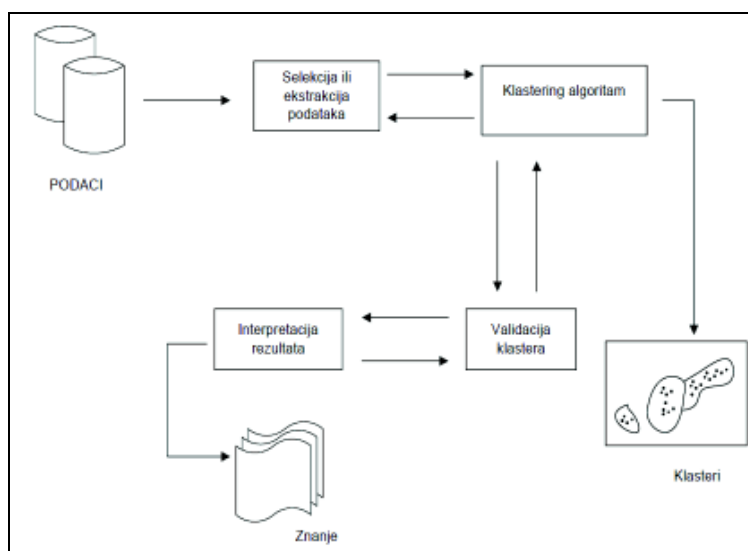
$$S = \bigcup_{i=1}^k C_i \quad i \quad C_i \cap C_j = \{\emptyset\} \quad \text{za } i \neq j$$

Posljedično, svaka instanca u S pripada jednom i samo jednom podskupu.

Dok se klaster definiše kao skupina sličnih objekata, klaster analiza ili klastering se definiše kao proces pronalaženja podskupova koji imaju zajedničke karakteristike. Klastering je deskriptivan proces, za razliku od većine metoda i tehnika koje su predikcijske. Mnogi se slažu da je to najraširenija metoda koja se koristi za razne vrste analiza. Proces klaster analize se obično dijeli u 4 osnovna koraka:

- Selekcija ili ekstrakcija podataka (uzoraka),
- Odabir klastering algoritma,
- Validacija klastera,
- Interpretacija rezultata.

Slika 1 prikazuje osnovne korake u procesu klaster analize.



Slika 1. Koraci klaster analize

Izvor: Nastavni materijali iz predmeta "Data mining" (Elektrotehnički fakultet Sarajevo)

2.1. Selekcija ili ekstrakcija podataka

Osnovni set podataka koji je korišten u ovom radu preuzet je iz DL sistema koji se već dugi niz godina koristi na Fakultetu informacijskih tehnologija u Mostaru. Podaci o uspjehu studenata su preuzeti, te za potrebe analize prilagođeni formatu koji podržava korišteni alat (Microsoft SQL Server Data Tools). Prilikom pripreme podataka za Microsoft klastering algoritam neophodno je osigurati sljedeće vrijednosti (kolone):

- Jedinstvena vrijednost (eng. Single key column),
- Ulazne vrijednosti (eng. Input columns),
- Vrijednost za predviđanje koja je opcionalna (eng. Optional predictable column).

U skladu sa mogućnostima korištenog alata, set podataka je prilagođen na sljedeći način:

- Jedinstvena vrijednost je identifikator studenta preuzet iz baze podataka,
- Ulazne vrijednosti su godina studija, način studiranja, status studenta, te prosjek ocjena po oblastima.
- Nije korištena vrijednost za predviđanje.

Primjer korištenog seta podataka prikazan je na slici 2 (na slici je prikazano prvih 10 zapisa iz korištenog seta podataka).

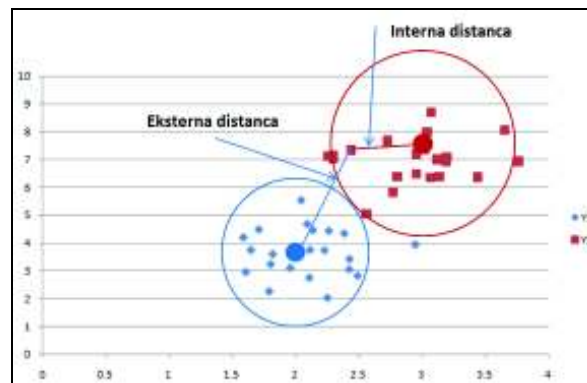
StudentID	GodinaStudija	NacinStudiranja	Status	ProsjekPRO	ProsjekSMA
1215	4	DL	Diplomirao	8	6
1217	4	DL	Diplomirao	8,25	8,67
1230	4	DL	Aktivan	7,78	6,75
1234	4	Redovni	Aktivan	7,18	6,5
1235	4	Redovni	Diplomirao	8,25	7,33
1238	4	Redovni	Aktivan	7	6,5
1241	4	DL	Diplomirao	7	7,33
1242	3	DL	Aktivan	7,57	7,2
1245	4	DL	Aktivan	6,25	8,33
1249	4	DL	Diplomirao	7,22	7,67

Slika 2. Dio korištenog seta podataka

2.2. Odabir klastering algoritma

Drugi korak u procesu klaster analize je odabir klastering algoritma. Algoritmi klaster analize teže da podijele skup objekata u podskupove u kojima je sličnost između objekata u jednom podskupu maksimizirana, a sličnost objekata između podskupova minimizirana. Ovaj korak se obično sastoji od određivanja odgovarajućih mjera distance. Uglavnom svi klastering algoritmi su eksplicitno ili implicitno povezani sa nekom od funkcija (mjera) distance. Kada se mjera distance odredi, klastering se može posmatrati kao optimizacijski problem u ovisnosti od te mjere. Postoje mnogi algoritmi za efikasnu klasterizaciju. U radu je korišten alat koji u svom jezgru posjeduje K-means klastering algoritam.

K-means je jednostavan iterativni algoritam koji skup podataka particionira u korisnički specificiran broj klastera. Naziv potiče od funkcije distance koja se primjenjuje – mean (eng. srednja vrijednost) i broja klastera koji se formiraju – k . Srednja vrijednost objekata u klasteru se uzima kao predstavnik klastera (centroid) i pripadnost objekta nekom klasteru se procjenjuje na osnovu distance od tog objekta. Na slici 3 je grafički predstavljen primjer K-means klasteringa.



Slika 3. K-means primjer klasteringa

Izvor: <https://www.jerrypost.com/Books/DMSummary/DataMiningChapter05.pdf> (12.5.2014)

Za mjeru distance objekata se najčešće uzima Euklidova distanca. Cilj klasteringa K-means algoritmom je da se minimizira vrijednost funkcije kvadratne greške:

$$E = \sum_{j=1}^k \sum_{x \in C_j} |x - m_j|^2$$

E je suma kvadrata grešaka za sve objekte u skupu podataka, a m_j je centroid klastera C_j kojem tačka x pripada. Kriterij za zaustavljanje algoritma može biti da je promjena greške E neutralizirana ili dovoljno mala vrijednost.

Za K-means klastering algoritam se obično kaže da je „hard“ algoritam. To znači da u procesu klasterizacije svaki objekat iz seta podataka može pripadati isključivo jednom klasteru, te da se vjerovatnoća pripadnosti nekom klasteru računa pojedinačno za svaki objekat. K-means se zbog jednostavnosti često koristi u hibridnim metodama gdje se kombinira sa drugim algoritmima kako bi se dobili bolji rezultati, povećala efikasnost algoritma, te smanjila kompleksnost.

2.3. Validacija klastera

U procesu klasteringa utvrđivanje broja klastera je veoma važan faktor. U ovisnosti od pripadnosti algoritma pojedinoj skupini postoje različiti principi određivanja broja klastera. Za mnoge klastering algoritme particioniranja, broj klastera k se zahtijeva kao korisnički specificiran parametar. Iako se u mnogim slučajevima k može procijeniti na osnovu korisničke ekspertize, također algoritam broj klastera može procijeniti iz podataka, bez da se unaprijed definiše. Prevelik ili premalen k utiče na kvalitet rezultirajućih klastera. Particioniranje sa mnogo klastera komplicira klastering strukturu čineći je teškom za interpretaciju i analizu rezultata. S druge strane, particioniranje sa samo nekoliko klastera uzrokuje gubitak informacija.

Kvalitet klasteringa je pojam koji nije moguće precizno definirati. Obično su značajni odgovori na pitanja koliko je dobar klastering generiran sa metodom, koliko dobro su klasteri odvojeni, koliko je kompaktan pojedini klaster. Za procjenu kvaliteta klasteringa često je potrebno porediti rezultate klasteringa dobivene sa različitim algoritmima, porediti strukturu i osobine dva klastera, ili dva seta klastera.

Na primjeru analize uspjeha studenata posebna pažnja je posvećena pravilnom odabiru relevantnih podataka, te transformaciji tih podataka u oblik prihvatljiv za odabrani algoritam. Na osnovu više testiranja i transformacija podataka u različite oblike, ustanovljeno je da u odabranom skupu podataka postoje klasteri (skupine studenata rangirane prema uspjehu), te da dalja analiza ima smisla.

Također, u postupku su korištene različite vrijednosti parametra koji definiše broj klastera, te različiti klastering algoritmi podržani odabranim alatom. Iz više pokušaja i dobivenih rezultata ustanovljeno je da K-means algoritam najbolje rasporedi studente u skupine rangirane prema uspjehu, i to ukoliko se unaprijed ne definiše broj klastera. Iako se za ovaj algoritam u literaturi često navodi da je



Larisa Dedović

Zvanje: Diplomirani inženjer informacijskih tehnologija

Tema diplomskog rada: „Unapređenje sistema za podršku učenju na daljinu – Studentski modul“, Datum odbrane: 30.09.2011. god.

2012 – danas: Master studij Fakulteta informacijskih tehnologija u Mostaru, Softverski inženjering

Projekti: QA Info Sys - Član tima za razvoj informacionog sistema za podršku internom osiguranju kvaliteta nastave na Univerzitetu „Džemal Bijedić“ u Mostaru, DLWMS v3 - Član tima za unapređenje i razvoj informacionog sistema za podršku učenju na daljinu, DLWMS v4 - Član tima za unapređenje i razvoj sistema za podršku učenju na daljinu (Distance Learning Web Management System version 4, UnSIS v5 - Član tima za unapređenje i razvoj informacionog sistema Univerziteta "Džemal Bijedić" u Mostaru (verzija 5)

Priznanja i nagrade: Rektorova nagrada Univerziteta "Džemal Bijedić", Bronzana plaketa za akademsku 2008/2009 za postignuti uspjeh 9.50 Rektorova nagrada, Univerzitet "Džemal Bijedić", Srebrna plaketa za akademsku 2009/2010 za postignuti uspjeh 9.73, Rektorova nagrada Univerziteta "Džemal Bijedić", Zlatna medalja za ostvareni uspjeh 9.36 tokom cjelokupnog studija, Nagrada Federalnog ministarstva obrazovanja BiH, Podrška za talentovane studente 2010.

unaprijed potrebno definisati broj klastera, alat podržava suprotno tj. u konkretnom primjeru postavljena je opcija da algoritam sam optimizuje i studente rasporedi u neodređeni broj klastera.

2.4. Rezultati klasteringa

Posljednji korak u procesu klasteringa je interpretacija rezultata. Skup klastera nije konačan rezultat i potrebno je detaljnije analizirati pojedine klastere i njihovo značenje. Kvalitet izvođenja prethodnih koraka direktno utiče na konačan ishod klasteringa. Nekada to mogu biti jako korisne informacije, a nekada beskorisne gomile podataka. U mnogim slučajevima klastering zahtijeva seriju pokušaja i ponavljanja. Za uspješnost analize bilo je neophodno filtrirati izvorni set podataka uzimajući u obzir način studiranja, godinu studija, te status studenta (aktivan, diplomirao). Zatim su za različite setove podataka analizirani dobiveni rezultati te njihova međuzavisnost. Postupak klasteringa izveden je nad podacima o uspjehu studenata iz dvije ključne IT oblasti (Softverski inženjering i Sistemska i mrežna administracija), primjenjujući sljedeće filtere:

- Svi studenti, nezavisno od načina studiranja, godine studija i statusa,
- Samo studenti starijih godina (3 i 4) koji su odslušali većinu predmeta iz pojedinih oblasti, nezavisno od načina studiranja i statusa,
- Studenti razdvojeni po načinu studiranja, nezavisno i zavisno od godine studija i statusa.

Da bi se dobili rezultati primjenjujući navedene filtere bilo je neophodno izvesti veliki broj pokušaja. U tabeli 1 izdvojeni su rezultati dobiveni iz 6 pokušaja, tj. prilagođavanjem izvornog seta podataka.

Pokušaj 1						
Svi studenti nezavisno od načina studiranja, godine studija i statusa						
<i>Variable</i>	<i>Klaster 1</i>	<i>Klaster 2</i>	<i>Klaster 3</i>			
Veličina klastera	881	65	57			
Prosjek SI	7,35 +/- 0,87	6,0	6,0			
Prosjek SMA	7,29 +/- 0,74	6,20 +/- 0,20	6,82 +/- 0,17			
Pokušaj 2						
Studenti 3 i 4 godine nezavisno od načina studiranja i statusa						
<i>Variable</i>	<i>Klaster 1</i>	<i>Klaster 2</i>	<i>Klaster 3</i>			
Veličina klastera	392	85	2			
Prosjek SI	7,66 +/- 0,81	6,69 +/- 0,26	6,75			
Prosjek SMA	7,63 +/- 0,67	6,60 +/- 0,28	7,78			
Pokušaj 3						
Diplomirani studenti nezavisno od načina studiranja						
<i>Variable</i>	<i>Klaster 1</i>	<i>Klaster 2</i>	<i>Klaster 3</i>	<i>Klaster 4</i>	<i>Klaster 5</i>	<i>Klaster 6</i>
Veličina klastera	48	31	20	16	4	3
Prosjek SI	7,04 +/- 0,32	8,33 +/- 0,45	7,24 +/- 0,20	9,20 +/- 0,37	7,21 +/- 0,24	8,3
Prosjek SMA	6,97 +/- 0,38	7,96 +/- 0,38	7,70 +/- 0,18	8,99 +/- 0,33	8,48 +/- 0,27	7,0
Pokušaj 4						
Studenti 3 i 4 godine uključujući način studiranja, nezavisno od statusa						
<i>Variable</i>	<i>Klaster 1</i>	<i>Klaster 2</i>	<i>Klaster 3</i>	<i>Klaster 4</i>		
Veličina klastera	237	193	42	7		
Način	R	0,30	0,40	0,38	0,0	

studiranja	DL	0,70	0,60	0,62	1,0
Prosjek SI		7,03 +/- 0,44	8,32 +/- 0,66	6,81 +/- 0,22	7,29
Prosjek SMA		6,94 +/- 0,39	8,09 +/- 0,67	7,83 +/- 0,25	8,0
Pokušaj 5					
Redovni studenti 3 i 4 godine nezavisno od statusa					
Varijable		Klaster 1	Klaster 2	Klaster 3	Klaster 4
Veličina klastera		79	23	21	28
Prosjek SI		7,09 +/- 0,47	8,66 +/- 0,33	9,19 +/- 0,50	7,40 +/- 0,29
Prosjek SMA		7,01 +/- 0,34	7,74 +/- 0,38	8,95 +/- 0,34	8,05 +/- 0,33
Pokušaj 6					
Distance learning (DL) studenti 3 i 4 godine nezavisno od statusa					
Varijable		Klaster 1	Klaster 2	Klaster 3	Klaster 4
Veličina klastera		277	38	12	2
Prosjek SI		7,60 +/- 0,72	6,63 +/- 0,18	6,45 +/- 0,22	7,31
Prosjek SMA		7,53 +/- 0,71	6,77 +/- 0,22	6,12 +/- 0,10	6,69

Tabela 1. Rezultati klasteringa²⁶⁶

3. ANALIZA REZULTATA

U većini pokušaja klasteringa posmatrani su samo studenti starijih godina koji su položili veći broj predmeta iz pojedinih oblasti, što je ključno za relevantnost podataka. Obzirom da je način studiranja jedan od kriterija za filtriranje podataka, kao i ulazna vrijednost u klastering algoritam, važno je naglasiti da se broj studenata u ovom kriteriju znatno razlikuje (70% DL i 30% redovni), što se odražava na rezultate.

U prvom pokušaju način studiranja je zanemaren, te je klastering izveden uzimajući u obzir samo uspjeh studenata tj. prosječnu ocjenu na pojedinim oblastima. Od ukupnog broja studenata čak njih 881 (87,83%) je svrstano u prvi klaster (*K1*). Budući da je prosječna ocjena na pojedinim oblastima jedini ulazni klastering parametar, ujedno je i osnovni kriterij za razmještaj studenata u klaster. Klaster 1 karakterišu studenti sa prosječnom ocjenom 7,66 (+/- 0,81) za oblast softverskog inženjeringa i 7,63 (+/-0,74) za oblast systemske i mrežne administracije, dok su u preostala dva klastera (*K2* i *K3*) smješteni studenti sa jako niskom prosječnom ocjenom iz obje oblasti. Uspjeh studenata u oblasti softverskog inženjeringa je zanemarivo bolji. Međutim, postavlja se pitanje koliko su rezultati pouzdani s obzirom da su u izvorni set podataka uključeni svi studenti bez posebnog filtriranja.

U drugom pokušaju izvorni set podataka je filtriran samo na studente 3 i 4 godine studija. Veličina seta podataka je znatno smanjena. U ovom slučaju relevantnost podataka je mnogo veća jer su u set uključeni studenti koji su položili većinu predmeta iz pojedinih oblasti. Dobiveni rezultati ne ukazuju na znatne promjene u odnosu na prvi pokušaj. Ponovo je najveći broj studenata smješten u prvi klaster (81,83%), te je vidljivo manje poboljšanje uspjeha u obje oblasti. Iz preostala dva klastera (*K2* i *K3*), također je primjetan napredak studenata po godinama.

²⁶⁶ SI – Softverski inženjering, SMA – Systemska i mrežna administracija
R – Redovni studenti, DL – Studetni koji studiraju po sistemu „Distance learning“

U trećem pokušaju posmatrani su samo diplomanti. Broj diplomiranih studenata u odnosu na cijeli set podataka je jako mali (12,16%). Klastering algoritam je proizveo 6 klastera, od čega prva tri (*K1*, *K2* i *K3*) čine većinu studenata i ponovo su to studenti sa prosječnim uspjehom u obje oblasti. Najzanimljiviji je četvrti klaster (*K4*) kojeg čine studenti sa izuzetnim uspjehom, gdje je prosječna ocjena iznad 9, te možemo reći da su to potencijalni IT stručnjaci. Budući da je riječ o diplomiranim studentima, zanimljivo bi bilo ispitati njihov status u kompanijama u kojima su zaposleni.

U četvrtom pokušaju kao ulazni parametar u klastering algoritam je uključen i način studiranja. U rezultatima nisu primjetna znatna odstupanja u odnosu na prethodne pokušaje. Najzanimljiviji je drugi klaster (*K2*) u kojem su smješteni uspješniji studenti. U ovom klasteru je primjetan porast broja redovnih studenata (40% u odnosu na ukupan broj), što ukazuje na to da redovni studenti imaju bolji prosjek.

To je upravo dovelo do naredna dva pokušaja (5 i 6), gdje je analiziran uspjeh studenata razdvajajući ih po načinu studiranja. Analizom rezultata uočene su sljedeće karakteristike vezane za redovne studente:

- Najveći broj redovnih studenata ima prosječnu ocjenu u obje oblasti (*K1*),
- Približno 50% studenata ima iznad prosječnu ocjenu u obje oblasti (*K2*, *K3*, *K4*)
- Prepoznat je klaster studenata sa izvanrednim uspjehom, čiji je prosjek $\approx 9,0$ (*K3*),
- Redovni studenti su bolji u oblasti softverskog inženjeringa (*K1*, *K2*, *K3*),
- Također, prepoznat je klaster u kojem su studenti bolji u oblasti systemske i mrežne administracije (*K4*).

Kada su u pitanju studenti koji studiraju po sistemu distance learning (DL), iz pokušaja 6 uočene su sljedeće karakteristike:

- Najveći broj DL studenata ima prosječnu ocjenu u obje oblasti (*K1*)
- Ostatak čine studenti sa lošim ocjenama, prosjek ispod 7,0 (*K2*, *K3*, *K4*).

4. DISKUSIJA I ZAKLJUČNA RAZMATRANJA

Dobivene rezultate neophodno je posmatrati sa stanovišta nastavnog plana i programa, organizacije i procesa nastave, te distribucije i kvalitete nastavnih materijala i sadržaja. Sve navedeno direktno utiče na uspjeh studenata, pa u konačnici i na kvalitet proizvedenog stručnog kadra.

Rezultati analize ukazuju na to da je konačan uspjeh studenata u manjoj prednosti u oblasti softverskog inženjeringa, posebno ako se uzmu u obzir redovni studenti koji prate nastavu na fakultetu. U tome se ogleda praktičan način izvođenja nastave po čemu je FIT prepoznatljiv. Također, analiza je pokazala da postoje redovni studenti sa izvanrednim uspjehom u obje oblasti koji se mogu okarakterisati kao potencijalni stručnjaci. Interesantno je da algoritam nije prepoznao klaster studenata koji su ostavili visoke ocjene u samo jednoj od oblasti. Broj ovakvih studenata sigurno bi bio još veći ukoliko bi prilikom upisa imali mogućnost odabira studijskog smjera.

Međutim, sa druge strane, DL studenti ne zaostaju mnogo s obzirom da nisu u mogućnosti da prate praktičnu redovnu nastavu u prostorijama fakulteta. Njihov uspjeh ogleda se u kvaliteti i distribuciji nastavnih materijala i sadržaja. Analiza je pokazala da je njihov konačan uspjeh u većini slučajeva prosječan. U posljednje vrijeme se posebna pažnja posvećuje ovom načinu studiranja, te se uvođenjem novih oblika nastave i distribucijom video/audio nastavnih sadržaja pokušava popularizirati i unaprijediti sistem distance learning-a. Posebno će biti

zanimljivo provesti sličnu analizu nakon određenog perioda, te uvidjeti uticaj noviteta na uspjeh DL studenata.

Rezultati analize treba da ukažu na kvalitet proizvedenog stručnog kadra što može biti iskorišteno u suradnji sa industrijom tj. potencijalnim poslodavcima. S ciljem unapređenja nastavnog plana i programa ovakav vid suradnje je izuzetno važan. Daljom analizom bi trebalo ispitati uspjeh na predmetima koji najviše doprinose profiliranju studenta u praktičnom pogledu, uključiti uspjeh iz srednjih škola te poslovni angažman diplomiranih studenata. Nakon određenog perioda, primjenjujući novi nastavni plan i program, bit će interesantno uraditi poređenje sa trenutnim stanjem.

REFERENCES

- [1] Nastavni materijali iz predmeta *Data mining*, Fakultet informacijskih tehnologija Mostar, 2014/2015 akademska godina (<https://student.fit.ba>)
- [2] Nastavni materijali iz predmeta *Data mining*, Elektrotehnički fakultet Sarajevo, 2012/2013 akademska godina (<http://c2.etf.unsa.ba>)
- [3] ZhaoHui Tang, Jamie MacLennan (2005) *Data mining with SQL Server 2005*, Wiley Publishing Inc. (10.5.2014)
- [4] Web, *Cluster Analysis*, <https://www.jerrypost.com> (12.5.2014)
- [5] Web, *Microsoft Clustering Algorithm*, <http://technet.microsoft.com>
- [6] Web, *Data Mining Clustering Example in SQL Server Analysis Services SSAS*, <http://www.mssqltips.com> (15.5.2014)
- [7] Web, *SQL Server Data mining Tutorial*, <http://georges.gardarin.free.fr> (15.5.2014)