

Analiza podataka u svrhu predlaganja studijskih programa u Nacionalnom informacijskom sustavu prijava na visoka učilišta

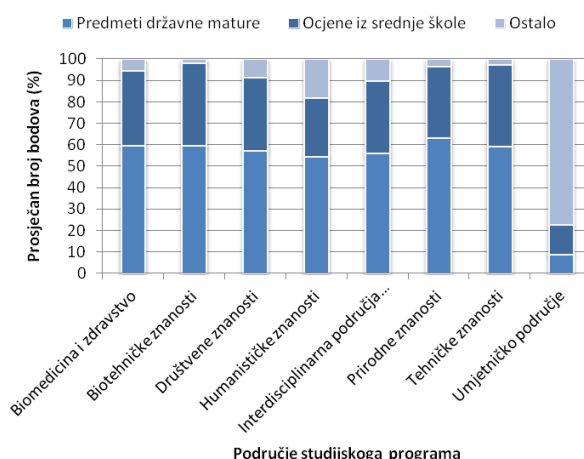
I. Budišćak*, V. Mornar* i D. Mlinarić*

* Fakultet elektrotehnike i računarstva, Sveučilište u Zagrebu, Hrvatska
ivan.budiscak@fer.hr, vedran.mornar@fer.hr, danijel.mlinaric@fer.hr

Sažetak – Nacionalni informacijski sustav prijava na visoka učilišta učenicima četvrtih razreda srednjih škola pruža mogućnost kreiranja svoje liste prioriteta najviše deset studijskih programa koje žele studirati. Neki od njih su neodlučni i ne znaju koje studijske programe odabrati. Učenicima se u sustavu prikazuje ostvareni broj bodova i plasman samo za studijske programe koji se nalaze na njihovoj listi prioriteta pa nisu u mogućnosti znati mogu li ostvariti pravo upisa na studijske programe koji nisu na njihovoj listi prioriteta. Budući da se rangiranje učenika obavlja periodički, svakih sat vremena, dodavanje i brisanje studijskih programa s liste prioriteta nije učinkovito rješenje za pronalazak studijskog programa koji odgovara učenikovim interesima i mogućnostima. U sklopu rada izradit će se model skladišta podataka za analizu ostvarenih rezultata učenika. Predložit će se postupak kojim će se neodlučnim učenicima predložiti studijski programi na temelju ocjena iz prva tri razreda srednje škole i postupak kojim će se odabrati studijski programi, koji se ne nalaze na listi prioriteta, za koje se učenicima dodatno može prikazati ostvareni broj bodova i mogući plasman.

I. UVOD

Za potrebe projekta državne mature razvili smo Nacionalni informacijski sustav prijava na visoka učilišta (NISpVU) koji objedinjuje cijeli postupak od prijave ispita državne mature i studijskih programa do rangiranja učenika za upis na studijske programe. Učenici mogu odabrati najviše deset studijskih programa koji čine njihovu listu prioriteta, a pravo upisa ostvarit će na onom studijskom programu koji se nalazi najviše na njihovoj listi prioriteta, a na čijoj se rang listi nalaze unutar upisne kvote. Slika 1. prikazuje prosječan udio bodova prema komponentama vrednovanja za sva područja studijskih programa. Najveći udio u bodovima, koji se u postupku rangiranja dodjeljuju učenicima, imaju rezultati ostvareni na ispitima državne mature i ocjene iz srednje škole. Kako bismo neodlučnim učenicima predložili studijske programe, uz ocjene koje su nam poznate za prva tri razreda srednje škole, potrebna nam je procjena rezultata ispita državne mature. Postupak za procjenu rezultata ispita državne mature opisat će se u III. poglavlju a u IV. poglavlju opisat će se postupak za predlaganje studijskih programa koji se temelji na prosječnoj ocjeni predmeta u prva tri razreda srednje škole i procijenjenom rezultatu ispita državne mature. U VI. poglavlju opisat će se model skladišta podataka iz kojega se mogu dohvatiti podatci potrebni za provedbu postupka predlaganja.



Slika 1. Prosječan udio bodova prema komponentama vrednovanja za sva područja studijskih programa

II. HRVATSKI SREDNJOŠKOLSKI SUSTAV [1]

Srednje škole, ovisno o vrsti nastavnog plana i programa koji provode, dijele se na:

- **gimnazije** - opće i specijalizirane.
- **strukovne škole** - tehničke, industrijske, obrtničke i druge, što se određuje prema vrsti nastavnog plana i programa.
- **umjetničke škole** - glazbene, plesne, likovne i druge, što se određuje prema vrsti nastavnog plana i programa.

Nastavni programi podijeljeni su u strukovna područja, a strukovna područja u obrazovne sektore.

Srednje obrazovanje učenika u gimnazijskim programima obrazovanja završava polaganjem državne mature (DM). Srednje obrazovanje učenika u strukovnim i umjetničkim programima obrazovanja, koji traju najmanje četiri godine, završava izradom i obranom završnoga rada u organizaciji i provedbi škole. Učenici četverogodišnjih strukovnih i umjetničkih srednjih škola mogu polagati i ispite državne mature, ako to žele.

Državna matura je skup ispita iz određenih nastavnih predmeta koje je učenik učio tijekom svoga najmanje četverogodišnjega srednjoškolskoga obrazovanja. Cilj je državne mature provjera i vrednovanje postignutoga

znanja i sposobnosti učenika, stečenih obrazovanjem prema propisanim općeobrazovnim nastavnim planovima i programima. Na državnoj maturi učenici prijavljuju tri obvezatna predmeta (Hrvatski jezik, Matematiku i strani jezik), a od ponuđena 22 izborna predmeta učenici mogu odabrati najviše šest [2].

Sva visoka učilišta u Republici Hrvatskoj prihvatila su rezultate ispita državne mature kao jedan od uvjeta za rangiranje učenika za upis na studijske programe. Osim vrednovanja rezultata ispita državne mature, visoka učilišta vrednuju i ocjene iz srednje škole, a za pojedine studijske programe uvjetuju se i dodatne provjere posebnih znanja, vještina i sposobnosti potrebnih za uspješno studiranje. Osim toga, kod upisa na mnoge studijske programe na sveučilištima u Republici Hrvatskoj vrednuju se i neka dodatna postignuća učenika. Zbroj bodova prema svim navedenim komponentama vrednovanja je broj bodova s kojima se na kraju postupka kandidati rangiraju za upis na visoka učilišta.

III. PROCJENA REZULTATA NA ISPITU DRŽAVNE MATURE

Procjenu rezultata na ispitu državne mature moguće je učiniti na temelju ostvarenih rezultata učenika iz prethodnih generacija. Možemo reći da, s određenom vjerojatnošću, učenici s boljom prosječnom ocjenom predmeta u prva tri razreda srednje škole ostvaruju bolji rezultat na ispitu državne mature [3]. Vjerojatnost je tim veća što je veći koeficijent korelacije između prosječne ocjene predmeta u prva tri razreda srednje škole i rezultata ostvarenoga na ispitu državne mature.

Za procjenu rezultata na ispitu državne mature, na temelju prosječne ocjene predmeta u prva tri razreda srednje škole, predlaže se korištenje postupka regresije [4], [5], [6]. U postupku regresije ciljna vrijednost $y \in \mathbf{R}$ je kontinuirana, a na temelju primjera $\mathbf{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}$ potrebno je naučiti nepoznatu funkciju $f: \mathbf{X} \rightarrow \mathbf{R}$ tako da, idealno, $y^{(i)} = f(\mathbf{x}^{(i)})$, gdje $\mathbf{X}$ predstavlja ulazni prostor primjera. Skup svih primjera $\mathbf{D}$ za postupak regresije, sastoji se od prosječne ocjene predmeta u prva tri razreda srednje škole i ostvarenoga rezultata (postotka riješenosti) na ispitu državne mature, za određeni skup učenika. Skup primjera može sadržavati podatke za učenike jednoga odjeljenja, jedne škole ili učenike koji pohađaju određeni program, strukovno područje ili obrazovni sektor.

Vrijednost koeficijenta korelacije između prosječne ocjene predmeta u prva tri razreda srednje škole i rezultata ostvarenoga na ispitu državne mature razlikuje se od škole do škole. Zbog toga je potrebno razviti više regresijskih modela, primjerice za svaku školu, umjesto jednog univerzalnog regresijskog modela za sve učenike.

IV. ODABIR STUDIJSKIH PROGRAMA ZA PREDLAGANJE

Odabir studijskih programa koji će se predložiti učeniku obavlja se na temelju procijenjenih rezultata ispita državne mature. Ovdje je bitno naglasiti da u izbornom dijelu učenici mogu prijaviti predmet koji nisu učili u srednjoj školi. To znači da za predmete koje učenik nije učio u srednjoj školi ne možemo napraviti procjenu rezultata na ispitu državne mature.

Postupak kojim će se učeniku predložiti studijski programi sastoji se od ovih sedam koraka:

- 1) Definira se skup svih studijskih programa $\mathbf{Q}$.
- 2) Odabire se skup učenika $\mathbf{U}$ čiji će se rezultati koristiti kao skup ulaznih primjera za učenje (50 %), provjeru (25 %) i testiranje (25 %) postupka regresije. To mogu biti učenici istoga odjeljenja, škole, programa, strukovnoga područja ili obrazovnoga sektora, koji su prethodne školske godine ostvarili upis na neki studijski program u ljetnome upisnom roku.
- 3) Odabire se skup predmeta $\mathbf{P}$. To su svi predmeti koje je učenik učio u srednjoj školi i za koje postoje rezultati ispita državne mature u prethodnoj školskoj godini za učenike iz skupa $\mathbf{U}$.
- 4) Za svaki predmet $p_i \in \mathbf{P}$, postupkom regresije opisanom u III. poglavlju, procjenjuje se učenikov rezultat na ispitu državne mature r_i .
- 5) Za svaki predmet $p_i \in \mathbf{P}$, odabire se skup učenika $\mathbf{K}_i \subseteq \mathbf{U}$ koji su na ispitu državne mature ostvarili rezultat u intervalu $[r_i - \varepsilon, r_i + \varepsilon]$. Na ovaj način obuhvatiti ćemo učenike koji su na ispitu ostvarili sličan rezultat, tj. rezultat koji je bolji ili lošiji za vrijednost ε . Vrijednost ε može se odrediti empirijski na podacima prethodnih generacija učenika.
- 6) Za svaki predmet $p_i \in \mathbf{P}$, svim studijskim programima $s_j \in \mathbf{Q}$ određuje se postotak učenika n_i iz skupa $\mathbf{K}_i$ koji su ostvarili pravo upisa.
- 7) Svakom studijskom programu $s_j \in \mathbf{Q}$ pridjeljuje se težina t_j određena medijanom vrijednosti iz skupa $\mathbf{M}_k = \{n_i\}$, $i \in [1, |\mathbf{P}|]$. Konačno rangiranje studijskih programa provodi se na temelju pridijeljene težine t_j .
- 8) Od tako rangiranih studijskih programa učeniku se predlaže prvih deset.

Umjesto promatranja studijskoga programa na kojemu su učenici ostvarili pravo upisa, u šestome koraku predloženoga postupka, možemo promatrati znanstveno polje ili područje studijskoga programa. Tako možemo dobiti rang-listu znanstvenih polja ili područja studijskih programa i dodatno prilagoditi predložene studijske programe. Za svako od najboljih deset znanstvenih polja ili područja studijskih programa, može se uzeti studijski program koji pripada tom znanstvenom polju ili području, a koji se nalazi najviše na rang-listi studijskih programa dobivenoj prethodno opisanim postupkom. Dodatnu prilagodbu predloženih studijskih programa može obaviti i učenik odabirom znanstvenih polja ili područja studijskih programa koji odgovaraju njegovim željama ili interesima ako ih ima.

Na temelju procijenjenih rezultata ispita državne mature i ostvarenih ocjena u prva tri razreda srednje škole, moguće je procijeniti i broj bodova za rangiranje koje će učenik ostvariti na ponuđenim studijskim programima. Usporedbom s rang-listama iz prethodnih godina, moguće je, s određenom vjerojatnošću, procijeniti učenikov plasman na ponuđenim studijskim programima a time i

hoće li ostvariti pravo upisa. Ako učenik želi studirati na nekom od ponuđenih studijskih programa na kojem s procijenjenim plasmanom ne ostvaruje pravo upisa, prikazana procjena može ga potaknuti na dodatno učenje predmeta koji se vrednuju za upis na tom studijskom programu. Dodatnim učenjem učenik bi mogao ostvariti bolje rezultate na ispitima državne mature a time i bolji plasman od procijenjenog.

V. ODABIR STUDIJSKIH PROGRAMA ZA DODATNI PRIKAZ OSTVARENIH BODOVA

Nakon što su poznati rezultati ispita državne mature u NISpVU postoje svi potrebni podatci za rangiranje učenika na svim studijskim programima. U tom trenutku promjene na rang-listama uzrokovane su isključivo izmjenama koje učenici rade na svojim listama prioriteta. Učenici koji nemaju pravo upisa na neke od željenih studijskih programa nemaju informaciju o plasmanu na srodnim studijskim programima ako se oni ne nalaze na njihovoj listi prioriteta. Ako učenik na svoju listu prioriteta doda studijski program, mora čekati izračun bodova koji se obavlja periodički, svakih sat vremena. Brisanje prijave za studijski program na kojemu je učenik ostvario pravo upisa, ima za posljedicu da učenik koji je na tom studijskom programu bio rangiran kao prvi koji se ne nalazi unutar upisne kvote ostvaruje pravo upisa a gubi ga na studijskom programu na kojemu je imao pravo upisa prije promjene. Promjene se nastavljaju lančano, dok se ne dođe do studijskoga programa kojemu je upisna kvota veća od broja prijavljenih učenika. Dodavanje studijskoga programa na kojem učenik po broju bodova ostvaruje pravo upisa ima za posljedicu gubitak prava upisa za učenika koji je bio rangiran kao posljednji unutar upisne kvote. Slično kao i kod brisanja prijave, promjene se nastavljaju lančano. Upravo zbog toga što prijava odnosno odjava može izazvati lančanu reakciju, izračun bodova obavlja se periodički.

Nakon svakoga izračuna bodova i generiranja rang-lista može se obaviti i izračun bodova za studijske programe koji se ne nalaze na učenikovo listi prioriteta. Izračun se može obaviti samo za one studijske programe koji pripadaju istom znanstvenom polju kao i studijski programi koji se na učenikovo listi prioriteta nalaze na višem prioritetu nego studijski program na kojemu je ostvareno pravo upisa. Dobiveni skup studijskih programa može se smanjiti ograničavanjem na studijske programe koji se izvode u županiji u kojoj učenik ima prebivalište ili prema nekomu drugom geografskom kriteriju. Alternativno se izračun može obaviti za studijske programe koje bi odredili postupkom opisanim u IV. poglavlju ali sada sa stvarnim rezultatima koje je učenik ostvario na ispitima državne mature.

VI. SKLADIŠTE PODATAKA

Za potrebe analize nužno je izraditi skladište podataka iz kojega se mogu iščitati potrebni podatci u prethodnim školskim godinama. Za izradu modela skladišta podataka korišten je Kimballov model [7]. Skladište podataka sastoji se od triju zvjezdastih shema, a sheme su povezane uporabom zajedničkih dimenzijskih tablica. Dimenzije i tablice činjenica opisat ćemo u nastavku ovoga poglavlja.

A. Dimenzije

Kreirane su ove dimenzije:

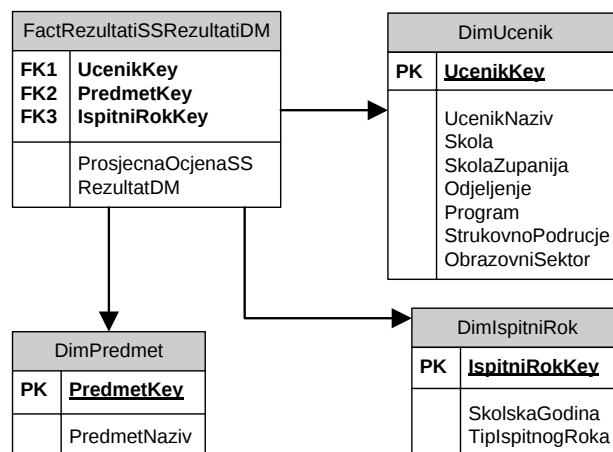
- *Učenik* – podatci o odjeljenju, školi, nastavnom programu, strukovnom području i obrazovnom sektoru
- *Predmet* – predmeti koji se polažu u sklopu državne mature
- *Studijski program* – podatci o nositelju i izvodaču, znanstvenom polju i području studijskoga programa
- *Razina ispita* – može biti viša ili osnovna
- *Ispitni rok* – školska godina i oznaka ispitnoga roka

U predloženim dimenzijama, dimenzija *Ispitni rok* može se smatrati vremenskom dimenzijom pri čemu je koristan pregled podataka prema školskim godinama ili ispitnim rokovima u jednoj školskoj godini.

B. Tablice činjenica

Slika 2. prikazuje tablicu činjenica koja se koristi kao izvor podataka za skup primjera za učenje, provjeru i testiranje postupka regresije opisanim u III. poglavlju. Uz tablicu činjenica prikazane su i pripadajuće joj dimenzijske tablice. Tablica činjenica sadrži ove atribute:

- *Učenik* – učenik koji polaže ispit državne mature
- *Predmet* – predmet koji se polaže ispitom državne mature
- *Ispitni rok* – ispitni rok na kojemu je proveden ispit državne mature
- *Prosječna ocjena* – prosječna ocjena u prva tri razreda srednje škole
- *Rezultat DM* – ostvareni rezultat na ispitu državne mature



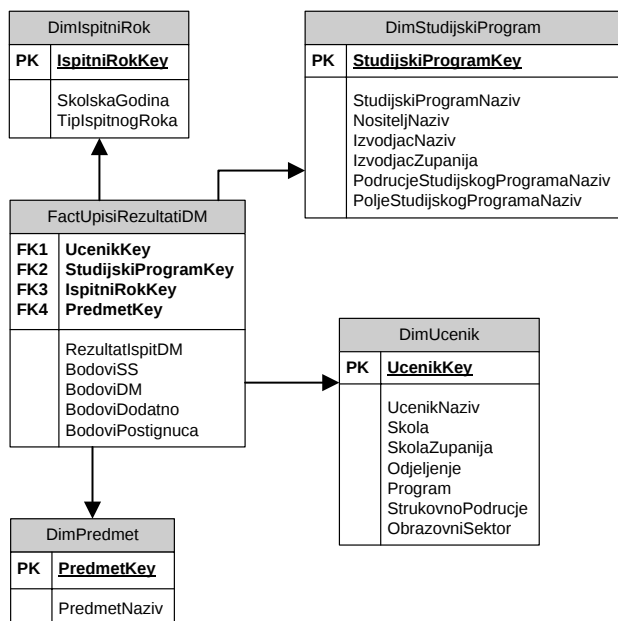
Slika 2. Tablica činjenica s pripadajućim dimenzijskim tablicama koja sadrži prosječnu ocjenu u prva tri razreda srednje škole i rezultat na ispitu državne mature za predmete koje su učenici polagali u sklopu državne mature

Podatci u tablici činjenica prikazanoj na slici 2. mogu se upotrijebiti i za određivanje korelacije između prosječnih ocjena predmeta u prva tri razreda srednje škole i rezultata ostvarenih na ispitima državne mature.

Pri izračunu rezultata ostvarenoga na ispitu državne mature valja naglasiti da se ispiti u obvezatnom dijelu državne mature polažu na dvjema razinama, osnovnoj i višoj [8]. Kako bi se rezultati ispita na osnovnoj i višoj razini mogli međusobno uspoređivati rezultat ostvaren na višoj razini množi se s faktorom 1,6 [8].

Slika 3. prikazuje tablicu činjenica koja se koristi u predloženom postupku za odabir studijskih programa za predlaganje, opisan u IV. poglavlju. Uz tablicu činjenica prikazane su i pripadajuće joj dimenzijske tablice. Tablica činjenica sadrži podatke o studijskim programima na kojima su učenici ostvarili pravo upisa i rezultate koje su učenici ostvarili na ispitima državne mature, a sadrži ove attribute:

- *Učenik* – učenik koji polaže ispit državne mature
- *Studijski program* – studijski program na kojemu je učenik ostvario pravo upisa
- *Ispitni rok* – ispitni rok na kojemu je proveden ispit državne mature i ostvareno pravo upisa na studijski program
- *Predmet* – predmet koji se polaže ispitom državne mature
- *Rezultat ispita DM* – rezultat ostvaren na ispitu državne mature
- *Bodovi DM* – broj bodova ostvaren na svim ispitima državne mature koji su vrednovani za upis na studijski program



Slika 3. Tablica činjenica s pripadajućim dimenzijskim tablicama koja sadrži podatke o studijskim programima na kojima su učenici ostvarili pravo upisa i rezultate na ispitima državna mature koje su ti učenici polagali

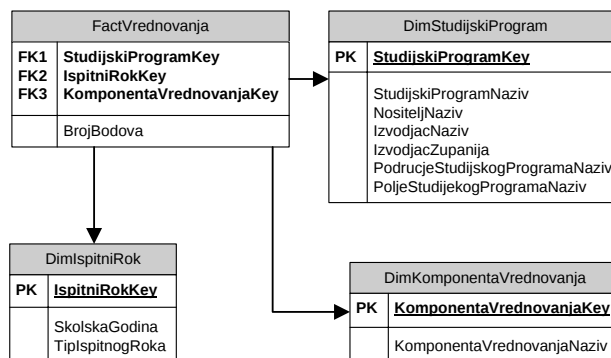
- *Bodovi SS* – broj bodova ostvaren temeljem ocjena iz srednje škole
- *Bodovi dodatno* – broj bodova ostvaren na provjerama posebnih znanja, vještina i sposobnosti
- *Bodovi postignuća* – bodovi ostvareni na temelju učenikovih dodatnih postignuća

Ovdje valja napomenuti da se prijave dvopredmetnih studija u NISpVU vode kao prijave dvaju studijskih programa pa je moguće da jedan učenik ostvari pravo upisa na dvama studijskim programima.

Iako se u predloženom postupku za odabir studijskih programa za predlaganje koristi samo rezultat ostvaren na pojedinačnim ispitima državne mature, tablica činjenica sadrži i sumarne bodove za studijski program na kojemu je učenik ostvario pravo upisa. Ti se bodovi mogu iskoristiti kako bi se dodatno ograničio skup studijskih programa koji će se predložiti učeniku. Primjerice možemo isključiti studijske programe na kojima su učenici posebnim postignućima i dodatnim provjerama ostvarili znatan broj bodova jer u tom slučaju bodovi ostvareni na ispitima državne mature nemaju jednaku težinu kao kod učenika koji te bodove nisu ostvarili.

Model podataka za opis vrednovanja studijskih programa je kompleksan i izvan okvira ovoga rada te ga ovdje nećemo opisivati. Bitno je naglasiti da se za svaku komponentu vrednovanja može odrediti maksimalan broj bodova koje učenici mogu ostvariti te da ukupan zbroj bodova iznosi 100. Slika 4. prikazuje tablicu činjenica korištenu za određivanje prosječnoga udjela bodova prema komponentama vrednovanja za sva područja studijskih programa. Uz tablicu činjenica prikazane su i pripadajuće joj dimenzijske tablice. Tablica činjenica sadrži podatke o maksimalnom broju bodova prema komponentama vrednovanja za sve studijske programe, a sadrži ove attribute:

- *Studijski program* – studijski program na kojemu je učenik ostvario pravo upisa
- *Ispitni rok* – ispitni rok za koji su definirana vrednovanja



Slika 4. Tablica činjenica s pripadajućim dimenzijskim tablicama koja sadrži maksimalan broj bodova prema komponentama vrednovanja za sve studijske programe.

- *Komponenta vrednovanja* – oznaka komponente vrednovanja
- *Broj bodova* – maksimalan broj bodova koji učenici mogu ostvariti za komponentu vrednovanja

VII. ZAKLJUČAK I PLANVI ZA DALJNJE ISTRAŽIVANJE

Kako bismo pomogli neodlučnim učenicima koji ne znaju koje studijske programe odabrati, u III. i IV. poglavlju predložen je jednostavan postupak za odabir studijskih programa koji im se mogu predložiti. Postupak se zasniva na procjeni rezultata ispita državne mature na temelju ocjena iz prva tri razreda srednje škole. Uspješnost predloženoga postupka može se provjeriti testiranjem na podacima u aktivnoj školskoj godini s time da se procjena temelji na podacima iz prethodne školske godine.

Predloženi postupak uzima u obzir samo jesu li ili nisu učenici upisali studijski program no ne i jesu li ga završili i s kojim uspjehom. Kada bismo imali podatke o uspjehu na upisanom studijskom programu, imali bismo i bolju informaciju o tome jesmo li predloženim postupkom odabrali studijske programe na kojima učenik može ostvariti najbolji rezultat, tj. ostvariti svoj potencijal. Analogno sustavima za predlaganje proizvoda za kupnju kao u [9], [10], gdje je potrebno uzeti u obzir i zadovoljstvo kupaca nekim proizvodom, ovdje se na sličan način može koristiti podatak o uspjehu na upisanom studijskom programu. Smatramo da bi se korištenjem tih podataka moglo pridonijeti profesionalnoj orijentaciji učenika koji upisuju studij. Analogno, postupak bi se

mogao upotrijebiti i za profesionalnu orijentaciju učenika koji upisuju srednju školu. Za takvu analizu uspjeh učenika trebalo bi se pratiti tijekom cijeloga obrazovanja, od osnovne škole do visokoga obrazovanja.

LITERATURA

- [1] Mrežne stranice Ministarstva znanosti, obrazovanja i sporta [Web]. <http://public.mzos.hr/Default.aspx?sec=2239>
- [2] Pravilnik o polaganju državne mature [Web]. <http://public.mzos.hr/fgs.axd?id=19419>
- [3] Statističko izvješće o rezultatima državne mature u školskoj godini 2009./2010. [Web]. http://dokumenti.ncvvo.hr/Dokumenti_centra/DM2010/Statistika/i_zvjestaj_2010.pdf
- [4] E. Alpaydm, Introduction to Machine Learning. The MIT Press, 2009, pp. 34-37
- [5] C.M. Bishop, Pattern Recognition and Machine Learning. Springer, 2006, pp. 138-177
- [6] S. Russell, P. Norvig, Artificial Intelligence: A Modern Approach (3rd Edition), Prentice Hall, 2009, pp. 717-727
- [7] R. Kimball, M. Ross, The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling. Wiley Computer Publishing, 2002.
- [8] Brošura za učenike u redovnom sustavu srednjeg obrazovanja u Republici Hrvatskoj [Web]. <http://public.mzos.hr/fgs.axd?id=19569>
- [9] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez. 2013. „Recommender systems survey“. Knowledge-Based Systems, vol. 46, pp. 109-132, July 2013.
- [10] Soltani, A.; Akbarzadeh-T, M.-R., "Confabulation based recommender system", Computer and Knowledge Engineering (ICCKE), 2013 3th International eConference on, vol., no., pp.82,86, Oct. 31 2013-Nov. 1 2013