Informacijska družba IS 2006 / Information Society IS 2006
9. mednarodna multi-konferenca / 9th International Multi-Conference
Vzgoja in izobraževanje v informacijski družbi / Education in Information Society
Ljubljana, Slovenija, 13.-14. oktober 2006

Napovedovanje učnega neuspeha

Prediction of Academic Failure

Dejan Rudolf
Srednja zdravstvena šola Ljubljana
dejan.rudolf@guest.arnes.si

Povzetek

Za učinkovito preprečevanje šolskega neuspeha potrebujejo šolski svetovalci orodje za napovedovanje učnega uspeha. Za takšne napovedi uporabljajo šolski svetovalci v slovenskih srednjih šolah predvsem ročno pridobljene podatke v povezavi s nekaterimi ugotovitvami in značilnostmi, ki jih odkrijejo v letih svojega dela. Predstavljena metoda bi lahko izboljšala zanesljivost teh napovedi.

Ključne besede: *napovedovanje učnega uspeha, izobraževalno svetovanje, odkrivanje znanja iz podatkov, odločitvena drevesa*

Abstract

For efficient prevention of academic failure counselors need a tool for prediction of final academic achievement. For such predictions counselors in Slovenian high schools use manually created data bases, based on expert knowledge. It is possibile that application and method proposed in this paper would increase the accuracy of predictions and quality of educational counseling.

Key words: *prediction of academic achievements, educational counselling, knowledge discovery from data, decision trees*

V slovenskih srednjih šolah, tudi strokovnih, se srečujemo z razmeroma visokim osipom dijakov med šolanjem. Razlogi so različni in segajo od napačne izbire šole oz. programa, pa do učnih in tudi osebnih težav, ki se pri posameznem dijaku pojavijo med šolanjem. Dijakom, pri katerih dovolj zgodaj odkrijemo težave lahko pomagamo z različnimi metodami, predvsem svetovanjem. Zato ima večina šol posebej zaposlene svetovalce. Težava, ki se pri tem pojavlja je, da svetovalci oz. svetovalne službe ne zmorejo opraviti celotnega dela, oz. ne uspejo detektirati dijakov, ki bi lahko imeli med šolanjem težave.

Pri tem so v pomoč učitelji in razredniki, pri katerih pa je to le ena od njihovih obveznosti in so marsikdaj prezaposleni, ob tem pa ne gre spregledati dejstva, da za tovrstno svetovanje pravzaprav niso usposobljeni.

Zdi se torej, da bi bile svetovalne službe mnogo bolj uspešne, če bi se že od začetka usmerile na prave dijake. Cilj je torej na nek način prečesati celotno populacijo in poskušati odkriti dijake, ki bodo potrebovali pomoč. S tem namenom poskušam zgraditi model, ki bo pokazal nekatere smernice oz. pravila, s pomočjo katerih bomo lahko s podatki, ki so nam na voljo ob začetku šolanja oz. ob koncu prvega letnika zaznali dijake s potencialnimi težavami in pravočasno ukrepali.

Eden glavnih razlogov za šolski neuspeh nedvomno tiči v dejstvu, da morajo učenci izbrati bodočo srednjo šolo in s tem v veliki meri tudi poklic, razmeroma mladi, večina pri starosti petnajst let. Ob vpisu je težko zagotoviti ustrezno svetovanje (Gasar, 2002), zato se prevečkrat zgodi, da učenci izberejo napačno šolo in to spoznajo šele takrat, ko so tja že vpisani. Pri takšnih dijakih pogosto opažamo nezanimanje za pouk, motenje pouka in izostajanje. Dogaja se tudi, da nekateri dijaki izberejo prezahteven program in so po prešolanju uspešni na enakem ali sorodnem področju. Dodatno na učni neuspeh vplivajo še:

- razmere v družini,
- (ne)pridobljene delovne navade,
- osebne težave,
- bolezen,
- neprimerno okolje itd.

Večine od teh parametrov ni mogoče enostavno zaznati neposredno, zato jh poskušamo zaznati posredno, preko količin, ki jih lahko merimo. Te poskušamo nato vključiti v ustrezen model, katerega napovedi lahko nato uporabimo pri delu.

Ob vpisu v srednjo šolo imamo načeloma na voljo nekatere podatke o učenčevem uspehu v osnovni šoli in nekatere osebne podatke, ki bi morda lahko bili pomembni. Med formalno pridobljenimi podatki so predvsem ocene in sicer:

- ocena iz slovenščine za zadnja tri leta OŠ,
- ocena iz matematike za zadnja tri leta OŠ,
- ocena iz tujega jezika za zadnja tri leta OŠ,
- splošni učni uspeh za zadnja tri leta OŠ.

Ti podatki za izdelavo modela niso zadostni, saj prihajajo dijaki iz različnih okolij in tudi iz različnih šol. Ocene so seveda do neke mere primerljive, popolnoma pa ne. Zato se zdi smiselno razmišljati v smeri pridobitve dodatnih podatkov. Do teh podatkov lahko pridemo šele ko so dijaki že vpisani na srednjo šolo oz. bolje rečeno med prvim letom šolanja na srednji šoli. Sem spadajo naslednje skupine podatkov:

- ocene prvega letnika
    - angleščina,
    - bologija,
    - fizika,
    - geografija,
    - informatika,
    - kemija,
    - matematika,
    - zdravstvena nega in prva pomoč,
    - slovenščina,
    - umetnost,
    - vzgoja za zdravje in
    - športna vzgoja
- demografski in ostali osebni podatki,

  - kraj bivanja,
  - število otrok v družini,
  - podatki o starših,
  - športne aktivnosti s katerimi se dijak ukvarja,
  - ostale aktivnosti,
  - udeležba pri prostovoljnem delu
- neformalni podatki
  - mnenje učiteljev in razrednika,
  - odnos dijaka do šole in šolskega dela,
  - udeležba pri interesnih dejavnostih,
  - pripravljenost na pomoč ipd.

Vse naštete podake pridobivamo v šolah na različne načine. Podatke iz prve skupine sestavljajo ocene in s pridobivanjem teh ni večjih težav, vsaj ne v tekočem šolskem letu. Prve težave se pojavijo že z ocenami iz osnovne šole. Za letnike, ki so se vpisovali od leta 2002 do 2005 so podatke na srednjih šolah dobili od centralne vpisne službe, ob vpisu dijakov. Za predhodne generacije podatkov v uporabni obliki ni, kot kaže pa jih glede na zadnje spremembe načina vpisa v srednje šole, tudi v bodoče ne bo.

Načeloma za generacije, ki jih imamo v šoli, podatke lahko pridobimo ročno oz. z anketami, vendar to predstavlja razmeroma velik projekt, saj moramo za takšno zbiranje podatkov pridobiti ustrezna soglasja (od staršev, pri polnoletnih dijakih pa od njih samih). To terja veliko pojasnevanja o namenu uporabe itd. Težave so že pri ocenah iz srednje šole, še veliko bolj pa so občutljiva zbiranja kakršnih koli drugih, predvsem osebnih podatkov.

Rezultat vsega naštetega je, da je težko pridobiti vse tiste podatke, ki bi si jih želeli oz. za katere menimo, da so pomemben pokazatelj za končni šolski uspeh ali neuspeh. Večino podatkov, ki predstavljajo ocene in opravičene ter neopravičene ure sem pridobil iz baze podatkov srednje šole v kateri sem zaposlen, s pomočjo programa Evidenca učencev, izdelek podjetja Pioma. Tem sem nato dodal še uspeh zaključnega letnika osnovne šole. Med pridobivanjem podatkov sem se srečeval s številnimi težavami, saj so bili pogosto podatki nepopolni ali pa tudi očitno napačni (npr. pri enem od predmetov je vpisana negativna ocena, končni uspeh prvega letnika pa je označen kot pozitiven). Vse takšne podatke je bilo potrebno pridobiti ali preveriti ročno, kar pa se je izkazalo za zelo zamudno.

Konči nabor je sestavljalo deset atributov, s pomočjo katerih sem nato raziskoval naprej. To so bili:

- splošni učni uspeh zaključnega razreda osnovne šole (OS8)
- ocena pri angleščini (ANJ)
- ocena pri matematiki (MAT)
- ocena pri strokovnem predmetu zdravstvena nega in prva pomoč (NPP)
- ocena pri slovenščini (SJK)
- splošni učni uspeh prvega letnika (US1)
- povprečje ocen prvega letnika (POV)
- število ur opravičenega izostanka (OPR)
- število ur neopravičenega izostanka (NEO)
- oceno ali je dijak pri šolanju uspešen ali ne (USPEH).

Razen splošnega učnega uspeha v zaključnem razredu osnovne šole in zadnjega atributa (USPEH) se vsi ostali nanašajo na prvi letnik srednje šole. Za zadnji atribut sem vrednosti določil tako, da sem preveril ali je v tekočem šolskem letu dijak vpisan v zaključni letnik (vsi podatki so za generacijo dijakov, ki se je vpisovala v šolskem letu 2002/2003). Skupno sem zajel podatke za 243 dijakov.

Model sem gradil s pomočjo orodja Weka, ki je brezplačen, javno dostopen program. Najdemo ga na spletni strani Univerze Waikato (http://www.cs.waikato.ac.nz/ml/weka/). Tam najdemo tudi nekaj navodil za uporabo programa, bolj izčrpne informacije pa so na voljo v obliki dokumentacije, generirane neposredno iz izvorne kode.

Orodje Weka (Waikato Environment for Knowledge Analyisis) omogoča uporabo več algoritmov s področja strojnega učenja. Sam sem uporabil dva in sicer Random Tree in algoritem imenovan J.48, oba uporabljata metodo odločitvenega drevesa. Analize za algoritem J.48 so podrobneje predstavljene v nadaljevanju.

Odločitvena drevesa so v splošnem razširjen način predstavitve znanja in se pogosto uporabljajo za reševanje klasifikacijskih problemov (Gasar, 2002). Odločitveno drevo je sestavljeno iz vozlišč, ki ustrezajo atributom, vej ki ustrezajo podmnožicam vrednosti atributov ter listov, ki ustrezajo razredom. Načeloma ne želimo imeti prevelikih dreves, ker so lahko zelo nepregledna. Prevelikim drevesom se izognemo z omejitvami (npr. števila elementov v listih) ali rezanjem.

Za primer lahko pogledamo izpis programa Weka in sicer odločitveno drevo, algoritem J.48 na desetih atributih in z rezanjem.

```
=== Run information ===

Scheme:       weka.classifiers.trees.J48 -R -N 3 -Q 2 -M 2
Relation:     uspeh
Instances:    243
Attributes:   10
              OS8
              ANJ
              MAT
              NPP
              SJK
              US1
              POV
              OPR
              NEO
              'USPEH'
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree
------------------
```

```
POV <= 2.36: NEUSPEH (11.0)
POV > 2.36
|   POV <= 3.29
|   |   US1 <= 1
|   |   |   OS8 <= 3: NEUSPEH (3.0)
|   |   |   OS8 > 3: USPEH (2.0/1.0)
|   |   US1 > 1: USPEH (69.0/10.0)
|   POV > 3.29: USPEH (77.0)

Number of Leaves  :     5

Size of the tree :      9


Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        217               89.3004 %
Incorrectly Classified Instances       26               10.6996 %
Kappa statistic                          0.4693
Mean absolute error                      0.1777
Root mean squared error                  0.3144
Relative absolute error                 66.8049 %
Root relative squared error             86.5425 %
Total Number of Instances              243

=== Detailed Accuracy By Class ===

TP Rate   FP Rate   Precision   Recall  F-Measure   Class
  0.99      0.632      0.894     0.99      0.94     USPEH
  0.368     0.01       0.875     0.368     0.519    NEUSPEH

=== Confusion Matrix ===

   a   b   <-- classified as
```

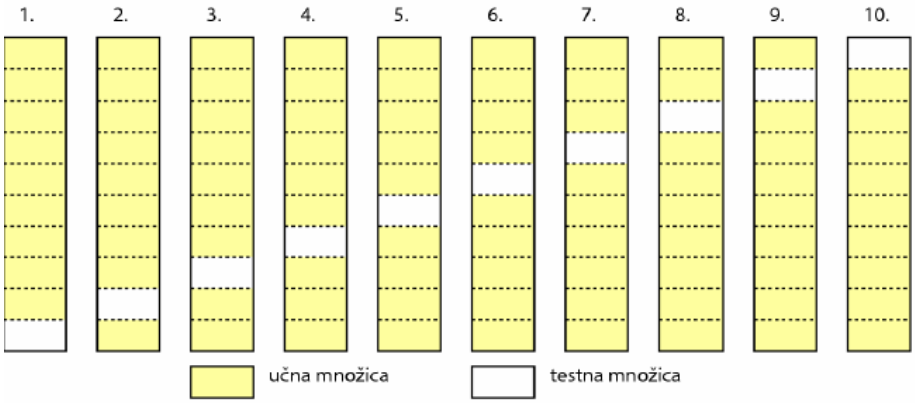
203 2 | a = USPEH

24 14 | b = NEUSPEH

Program je klasificiral podatke glede na algoritem, ki smo ga izbrali. Gre za algoritem J.48, ki ga uporabljajo številni programi. Vidimo lahko, da gre za primer odločitvenega drevesa v razmeroma omejenem obsegu. Tako oklesten še ni dovolj za ustrezne napovedi in uporabo v praksi, vendar ga je smiselno vseeno analizirati, saj je celotno drevo, z vsemi potrebni parametri, za tukajšnjo predstavitev preobširno, samo delo pa poteka po enakem vzorcu. S spremembo parametrov dosežemo ustrezno klasifikacijo in s tem tudi pregledno drevo. Ne glede na omejen obseg, lahko tudi na danem primeru razberemo način delovanja programa Weka. Kot glavni klasifikator je algoritem na danem vzorcu izbral povprečno oceno v prvem letniku. Ugotovimo, da so bili vsi dijaki s povprečjem pod 2.36 neuspešni in da je takšnih dijakov 11. Nato klasificira ponovno in sicer še enkrat po povprečju. Tokrat je meja na 3.29. Naslednji atribut je učni uspeh v osmem razredu osnovne šole. Vsakič lahko na izpisu tudi preberemo koliko dijakov je padlo v določen razred.

Po izgradnji modela je smiselno napovedi tudi preveriti. Osnovni postopek je, da množico podatkov razdelimo na dve množici in sicer učno in testno (npr v razmerju 2:1). Pri tem opazujemo, kako je model, ki smo ga razvili na učni množici, uspel klasificirati nove primere iz testne množice.

Druga možnost pa je, da uporabimo večkratno prečno preverjanje. Postopek je podoben prej opisanemu. En del množice izločimo in na preostanku izgradimo model, ki ga uporabimo za razvrščanje izločenega dela podatkov. Na enak način nato izločimo tudi vse ostale dele in jih preverimo (Robič, 2005).



Slika 1: prikaz 10-kratnega prečnega preverjanja

Na ta način preverja klasifikacijsko točnost tudi program WEKA, ki na koncu izpisa tudi predstavi, koliko elementov je s pomočjo uporabljenega modela klasificiral pravilno in koliko ne.

Dosežene natančnosti, določene na opisan način, so za obravnavani primer znašale med 80 in 90 %, kar kaže na to, da bi se kazalo z modelom še ukvarjati. Tu mislim tako na izbiro podatkov, ki jih vključujemo v model, kakor tudi na izgradnjo samega modela in nastavitve v njem.

Zaradi sprememb v družbi in posledično spremenjenega položaja šol, smo tudi v slovenskih srednjih šolah začeli razmišljati drugače, kot je bila praksa do sedaj. Ni več vseeno kakšen je uspeh dijakov, ki obiskujejo posamezno šolo. Vse več je prizadevanj, da bi se dvignila kvaliteta storitev, ki jih šole ponujajo. Slednje se nanaša tako na sam pedagoški proces, kakor tudi na podporne aktivnosti. Tako se na šolah pojavlja tudi vse več projektov, usmerjenih v reševanje učne problematike. Ena od možnosti, kako izboljšati tovrstne projekte, je tudi pričujoči model napovedovanja učne (ne)uspešnosti.

Na Srednji zdravstveni šoli Ljubljana uporabljam predstavljeni model testno na dveh oddelkih (59 dijakov). Zaenkrat so bili (do tretjega letnika) neuspešni trije dijaki za katere je model to tudi

predvidel, končno neuspešnost (pred maturo) pa napoveduje še štirim do šestim dijakom (odvisno od nastavitve parametrov). Tem dijakom poskušam kot razrednik in učitelj posvečati večjo pozornost. Nanje opozarjam tudi svetovalno službo s katero tesno sodelujem. Napovedi modela so skladne z napovedmi eksperta. Ob tem poskušam model tudi nadgrajevati z novimi atributi (ocene višjih letnikov, opažanja razrednika in učiteljev med šolanjem). Pri tem sem zelo pazljiv, saj menim, da dijaki ne smejo vedeti, za napovedi modela, saj bi to lahko posredno vplivalo na njihovo uspešnost. Glavna težava, da modela ni mogoče enostavno razširiti na celotno šolo, pri nas, po mojem mnenju tiči v neustrezno organizirani in urejeni bazi podatkov. Menim, da bo prav skrbi za kvaliteto podatkov, v prihodnje potrebno posvečati več pozornosti.

Literatura in viri

Gasar Silvana, Večparametrski model za predvidevanje uspešnosti zaključka šolanja, magistrsko delo, Fakulteta za organizacijske vede, Maribor, 2002

Robič Tea, Strojno učenje s programom Weka, IJS, Ljubljana, 2005

Weka, Univerza Waikato, (http://www.cs.waikato.ac.nz/ml/weka/), 2005

Strokovni življenjepis

Dejan Rudolf *je leta 1998 diplomiral na pedagoški smeri Oddelka za fiziko, Fakultete za matematiko in fiziko, Univerze v Ljubljani. Na isti fakulteti je leta 2003 zaključil tudi dodatno izobraževanje za učitelje informatike. Trenutno je slušatelj podiplomskega, magistrskega študija poslovne informatike, na Ekonomski fakulteti v Ljubljani, pri čemer se največ ukvarja s poslovnimi procesi, informatizacijo in sistemi za podporo odločanju.*

Na Srednji zdravstveni šoli Ljubljana, kjer je zaposlen, je poleg poučevanja fizike in kasneje tudi informatike, opravljal številne dodatne zadolžitve, kot so vodenje študijske skupine v okviru Centra za poklicno izobraževanje RS, vodenje strokovnega aktiva, izgradnje šolskega informacijskega sistema ter vodenje številnih aktivnosti na šoli povezanih z informatiko. Je član strokovne komisije za prenovo srednješolskih programov iz fizike pri Zavodu za šolstvo RS in član sveta Srednje zdravstvene šole Ljubljana.