

## PROVERA ZNANJA PRIMENOM RAČUNARSKOG ADAPTIVNOG TESTIRANJA KAO ODGOVOR NA SAVREMENE TENDENCIJE U OBRAZOVANJU

Svetlana Anđelić<sup>1,2</sup>, Nikola Dragović<sup>2</sup>, Goran Radić<sup>2</sup>, Valentin Kuleto<sup>2</sup>, Velimir Dedić<sup>1</sup>

<sup>1</sup>Univerzitet PIM Banja Luka, Računarski fakultet, despota Stefana Lazrevića bb, 78000  
Banja Luka, Bosna i Hercegovina

<sup>2</sup>Visoka škola informacionih tehnologija ITS- Beograd, Savski nasip 7, 11000 Beograd, Srbija,  
svetlana.andjelic@its.edu.rs

### SAŽETAK

Obrazovanje, na svim nivoima, je pretrpelo određene izmene usled pandemije virusom Covid-19 – nastava se održavala isključivo putem online platformi ili uz eventualnu kombinaciju tradicionalne, frontalne nastave. Pojedine ustanove su parcijalne provere znanja takođe obavljale putem online platformi uz eventualno uključivanje video call meetinga. Promene koje su tada usvojene su dobrim delom zadržane i nakon prestanka same pandemije.

Online testovi za samoprocenu i procenu znanja se uglavnom zasnivaju na random generisanju pitanja i odgovora. Ovakvi testovi nisu prilagođeni trenutnom znanju samih ispitanika i samim tim negativno utiču na njihovu motivaciju. Računarsko adaptivno testiranje omogućava da se test, odnosno generisana pitanja postavljaju u zavisnosti od ispitanikovog odgovora na prethodno postavljeno pitanje (simulacija usmenog ispitivanja). Ovaj način procene znanja ima dvojake pozitivne efekte, kod ispitanika stvara utisak da je test kreiran "samo za njega", dok sa druge strane povećava efikasnost testiranja jer se smanjuje broj pitanja na koje ispitanik treba da odgovori kako bi se objektivno moglo proceniti njegovo znanje.

U ovom radu će biti objašnjen koncept originalnog, autorskog CAT modela testiranja. Biće prikazani i rezultati testiranja u visokom i srednješkolskom sistemu obrazovanja koji su sprovedeni primenom navedenog CAT modela.

**Ključne reči:** obrazovanje, online test, CAT model, studenti, učenici.

### UVOD

Cilj svake ocene, odnosno ispitivanja studenata i učenika je da se što objektivnije utvrdi stepen odnosno nivo njihovog znanja iz određene nastavne materije. Pojam objektivnosti je često upitan, posebno kod odgovora esejskog tipa, jer nastavnik samostalno daje poene na osnovu sopstvenog kriterijuma i utiska sveohvatnosti napisanog odgovora. Dalji problem je i u broju postavljenih pitanja i potpitanja kako bi se obuhvatio što širi opseg gradiva, kao i u kompleksnosti tih pitanja. Za veću ocenu se očekuje da ispitanik (student, učenik) zna odgovore na složena i zahtevna pitanja koja povezuju više lekcija i/ili oblasti. Ako se svemu ovome doda veliki broj studenata/učenika koje nastavnik treba da ispita u zadatom vremenskom roku dolazi se do zaključka da klasično ispitivanje (pismeni u papirnoj formi i usmeno ispitivanje) nije baš efikasno sredstvo (vreme ispitivanja + vreme pregledanja radova + mogućnost ljudske greške).

### RAČUNARSKO ADAPTIVNO TESTIRANJE

Savremene informaciono komunikacione tehnologije se uveliko koriste u obrazovne svrhe i online testiranje se sprovodi na svim nivoima obrazovanja. Primenom računarskog testiranja je smanjen stepen subjektivizma nastavnika i svakako otklonjena mogućnost ljudske greške pri pregledanju testova. Većina ovih testova se zasniva na generisanju pitanja i odgovora po principu slučajnog (eng. *random*) izbora pitanja, odnosno random promešanog redosleda ponuđenih odgovora. Problem ovako kreiranih testova je u tome što svi ispitanici dobijaju ista pitanja bez

obzira na nivo znanja koji poseduju. Naime, kod ispitanika koji imaju visok nivo znanja (odgovaraju za veću ocenu) apsurdno je tražiti da odgovaraju i na trivijalna pitanja i tako gube koncentraciju i produžuju vreme trajanja ispitivanja. S druge strane, kod ispitanika koji imaju znatno niži nivo znanja i odgovaraju za niže ocene ne treba raspasti pažnju na previše zahtevna i složena pitanja.

U suštini, problem ispitivanja treba sprovesti na što efektivniji i efikasniji način. Maksimalna informacija se dobija kada se ispitaniku postavljaju pitanja koja odogovaraju njegovom stepenu znanja (Anđelić, 2010). Odnosno, uz što manji broj postavljenih pitanja (a samim tim i potrebnim vremenom za odgovaranje) utvrditi što objektivnije nivo znanja ispitanika.

Upravo to je rešeno primenom računarskog adaptivnog testiranja (eng. *Computer Adaptive Testing – CAT*). Adaptivnost se ogleda u tome što naredno pitanje (koje se generiše po specifičnom algoritmu) zavisi od tačnosti odgovora na prethodno pitanje. Takav način individualnog testiranja omogućava da vrlo mali broj ispitanika rešava identični test (Anđelić, 2010). Ova vrsta testiranja je specifična jer se ne može sprovesti primenom papirnog testa.

CAT proces se sastoji iz petlje koja procenjuje nivo studentovog znanja na osnovu odgovora na prethodna pitanja i traži pitanje koje će dati najviše informacija saglasno nivou ispitanikovog znanja (Anđelić, 2010).

U svetu postoji više različitih modela koji izračunavaju (predviđaju) verovatnoću da ispitanik tačno tj. netačno odgovori na postavljeno pitanje i na osnovu tih verovatnoća se vrši procena nivoa znanja ispitanika. Neki od značajnijih modela su:

Teorija davanja odgovora (eng. *Item Response Theory, IRT*) – kod ovog sistema se procena nivoa znanja ispitanika bazira na podacima dobijenim predtestiranjem uzorka ispitanika (Baker, & Ho Kim, 2017); glavni nedostatak je što pitanja nije moguće podeliti po određenim kategorijama i to dovodi do znatnog povećanja vremena pretrage kako se baza pitanja povećava.

Sekvencijalni test odnosa verovatnoća (eng. *Sequential Probability Ratio Test, SPRT*) – za razliku od IRT modela kod ovog modela nema predtestiranja i implementacija je znatno jednostavnija (Stefan, Schönbrodt, Evans, & Wagenmakers, 2022); primer primene je testiranje za dobijanje TOEFL (eng. *Test of English as a Foreign Language*) sertifikat poznavanja engleskog jezika za strance.

Bajesove mreže (eng. *Bayesian Networks, BN*) – glavna prednost ovog modela je mogućnost organizovanja po težinskim nivoima te se do procene nivoa znanja ispitanika dolazi nakon par iteracija; nedostatak se ogleda u tome što je za dobijanje početnih parametara neophodno izvršiti predtestiranje na velikom uzorku ispitanika (Scutari, & Denis, 2021).

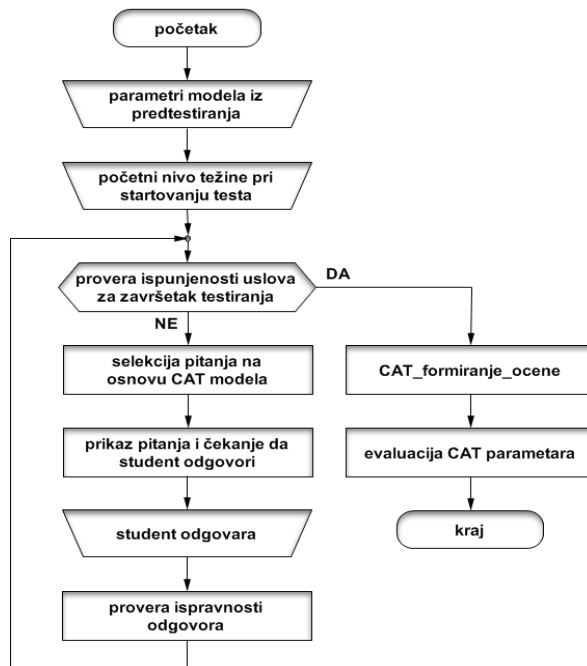
Parcijalni red strukture znanja (eng. *Partial Order Knowledge Structures, POKS*) – zasniva se na BN modelu pri čemu ne zahteva veliki uzorak za predtestiranje (Beheshti, & Desmarais, 2012).

Svi ovi modeli dolaze do sličnih rezultata u proceni znanja, uzimajući u obzir prosečne vrednosti testiranih ispitanika, a ne pojedinačno za svakog ispitanika ponaosob (Desmarais, & Pu, 2008).

## **SPECIJALIZOVANI CAT MODEL**

Sagledavajući osnovne nedostatke ranije pomenutih modela razvijen je specijalizovani CAT model koji nastoji da prevaziđe te nedostatke. Ovaj model je autorski rad (kroz doktorsku disertaciju) prvog autora ovog rada. U nastavku će ukratko biti prikazna taj model, a glavni akcenat će biti na primeni tog modela pri testiranju uzoraka ispitanika u srednjem i visokom obrazovanju.

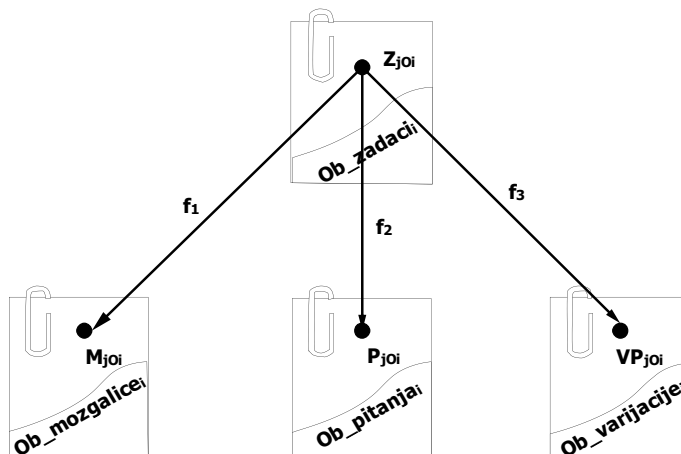
Sam CAT model se može prikazati kao iterativni algoritam (slika 1) na kome se jasno uočava da se nakon unosa neophodnih parametara iz predtestiranja ti parametri nadalje sami evaluiraju, odnosno nakon svakog testiranja ispitanika parametri se automatski ponovo preračunavaju i tako korigovani parametri predstavljaju ulazne parametre za naredno testiranje.



Slika 1. Proces specijalizovanog CAT sa detaljnijim prikazom CAT testiranja (Anđelić, 2010).  
 Figure 1. Process of specialized CAT with a more detailed view of CAT testing (Anđelić, 2010).

Odnosno, ovo ukazuje da što više puta se sistem pokrene to su parametri sve precizniji i daju bolju predikciju u selekciji pitanja koja se ispitaniku postavljaju. Bitno je naglasiti (što se može i videti u prikazanom algoritmu) da su selekcija pitanja i generisanje ocena dva nezavisna modula koja se mogu menjati bez značajnog uticaja jednog na drugi.

Svaka oblast ima definisana tri težinska nivoa (najviši, srednji, osnovni). Na najvišem nivou su zadaci i mozgalice (teži zadaci), na srednjem nivou su pitanja i varijacije pitanja (pitanje sa datim pojašnjenjem), a na najnižem nivou su osnovna (bazna) pitanja. Nastavnik definiše zadatke za svaku oblast, kao i koje oblasti učestvuju u testiranju. Zadaci se jednoznačno preslikavaju (slika 2) u mozgalice pitanja i varijacije pitanja.



Slika 2. Funkcije preslikavanja između zadataka, mozgalica, pitanja i varijacija (Anđelić, 2010).  
 Figure 2. Mapping functions between tasks, puzzles, questions and variations (Anđelić, 2010).

Ispitanik sam bira nivo sa kojim započinje testiranje, odnosno nivo od koga startuje u odgovaranju prve oblasti. Svaka sledeća oblast započinje od nivoa kojim je završena prethodna oblast. Završetak testiranja u okviru jedne oblasti se dešava kada se ispitanik u najmanje dva (ovaj broj se može menjati po potrebi) postavljena zadatka (ili njegove funkcije preslikavanja prikazane na slici 2) budu na istom težinskom nivou. Ispitanik prolazi kroz sve oblasti koje je nastavnik definisao za dati test.

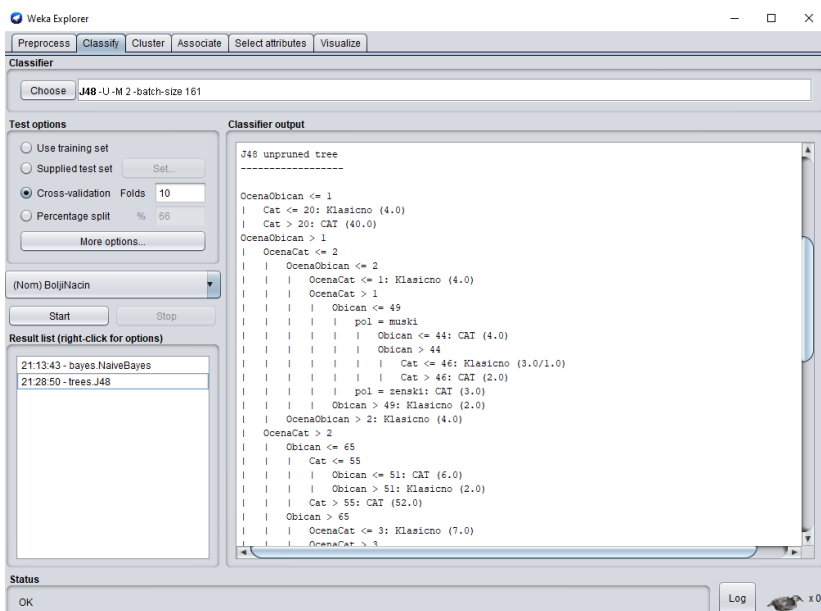
Specijalizovan algoritam za generisanje ocena je definisan tako da se izračunava aritmetička sredina ponderisane ocene (na osnovu sakupljenog broja poena) i ocene dobijene na osnovu Bajesove teoreme uslovne verovatnoće i Maksimum aposteriori pristupa. Odnosno, nakon svakog pitanja na testu izračunaju se poslednje aposteriori verovatnoće i određuje se maksimalna vrednost među njima. Ocena za koju je utvrđeno da njena verovatnoća odgovara izračunatoj maksimalnoj vrednosti je tražena „najverovatnija” ocena (Anđelić, 2010).

## REZULTATI EKSPERIMENTA

Eksperiment je sproveden u okviru dva nivoa obrazovanja, visokom i srednješkolskom.

U visokom obrazovanju testirano je 100 studenata primenom specijalizovanog CAT modela i klasičnim putem. Kod klasičnog testiranja su svi studenti odgovarali na istih 14 pitanja iz tri oblasti. Kod CAT testiranja prosečno je odgovarano na 5.6 pitanja (maksimalno na 11 pitanja i minimalno na 3). Poređenjem rezultata i primenom Mann-Whitney testa i Spirmanovog koeficijent korelacije utvrđeno je da je poklapanje ocena bilo kod 82.57% slučajeva. Odatle sledi jasan zaključak da je broj pitanja, a samim tim i vreme testiranja kod CAT modela znatno manji a dobijeni rezultati se u znatnoj meri poklapaju.

Eksperiment je sproveden na uzorku od 161 učenika srednje škole u više razreda. Za pronalaženje povezanosti između podataka korišćena je WEKA softver (eng. *Waikato Environment for Knowledge Analysis*) koji se sastoji od razlikih algoritama za mašinsko učenje. Primenjivana su dva klasifikatora, J48 (slika 3) i Naïve Bayes klasifikator (slika 4) (slično kao Hussain, Dahan, Ba-Alwib, & Ribata, 2018). Primenom oba klasifikatora se došlo do sličnih rezultata (poklapanje u 93.78% kod J48 i 86.33% kod Naïve Bayes). Ovo se poklapa sa istraživanjima drugih autora koji su utvrdili da J48 klasifikator daje bolje rezultate (Sumitha, & Vinothkumar 2016).



Slika 3. Pregled rezultata dobijenih izvršavanjem klasifikatora J48.  
Figure 3. Overview of the results obtained by running the J48 classifier.

```
Time taken to build model: 0.01 seconds

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

Correctly Classified Instances      139          86.3354 %
Incorrectly Classified Instances    22           13.6646 %
Kappa statistic                    0.4505
Mean absolute error                 0.3399
Root mean squared error             0.3907
Relative absolute error             111.0948 %
Root relative squared error         100.3351 %
Total Number of Instances          161

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area
                0.400   0.031   0.750     0.400   0.522     0.481   0.676   0.562
                0.969   0.600   0.876     0.969   0.920     0.481   0.676   0.840
Weighted Avg.   0.863   0.494   0.852     0.863   0.846     0.481   0.676   0.788

=== Confusion Matrix ===

  a  b  <-- classified as
12 18 | a = Klasicno
 4 127 | b = CAT
```

Slika 4. Pregled rezultata dobijenih izvršavanjem Naïve Bayes klasifikatora.  
Figure 4. Overview of the results obtained by running the Naïve Bayes classifier.

## NEDOSTACI CAT TESTIRANJA

Kao i kod većine drugih CAT modela (mada se ovo odnosi na svako testiranje primenom računara) glavni nedostatak je taj što se oni ne mogu primeniti na sve teme i veštine (npr. slikanje, čitanje,...).

Takođe, kao eventualni nedostatak se može navesti nemogućnost povratka na prethodno pitanje jer se svako sledeće pitanje postavlja kao rezultat tačnosti odgovora na prethodno pitanje. Ali, ako se ovo posmatra sa druge strane uvideće se da je ovo prednost a ne mana. Naime, ako bi se omogućio povratak na prethodno pitanje, onda bi se mogla dogoditi situacija da “mudriji” ispitanici kad shvate da je sledeće pitanje lakše shvate da su dali pogrešan odgovor na prethodno pitanje i da onda idu principom “pogađanja” i ispravljanja odgovora, što bi dovelo do stravanja lažne slike o nivou ispitanikovog znanja.

Prilikom uvođenja prikazanog CAT modela može se očekivati određeni otpor među nastavnicima jer se od njih očekuje da definišu veliki broj zadataka sa njihovim pratećim mozgalicama, teorijskim pitanjima, varijacijama tih pitanja i osnovnim pitanjima, kao i 5 varijanti odgovora (1 tačan i 4 distraktora, kako bi se ispoštovala sva pravila pri kreiranju testa kao objektivnog mernog instrumenta (Nedelesky, 1954; Kehoe, 1995).

Ukoliko je komplikovano i/ili nemoguće izvršiti pretestiranje na osnovu kojih bi se odredili svi neophodni ulazni parametri modela (apriori verovatnoće svih pitanja) mogu se uzeti identične verovatnoće za sve ocene (kod studenata su ocene u rasponu od 5-10, odnosno apriori verovatnoća je 1/6, dok je u srednjoj školi raspon ocena od 1-5, pa bi apriori verovatnoća bila 1/5).

## ZAKLJUČCI

Prikazani model i aplikacija CAT testiranja su fleksibilni.

Na osnovu dobijenih rezultata eksperimenta može se zaključiti da CAT testiranje ima dvojake pozitivne efekte. S jedne strane, kod ispitanika se stvara utisak da je test kreiran “samo za njega”, a sa druge strane povećava se efikasnost testiranja jer se smanjuje broj pitanja na koje ispitanik treba da odgovori kako bi se objektivno moglo proceniti njegovo znanje.

Specijalizovani CAT model bi značajnu ulogu mogao da ima kao pomoćno didaktičko sredstvo u svim nivoima obrazovanja, a posebno sa aspekta samoprocene znanja učenika i/ili studenata u toku nastavnog procesa.

## LITERATURA

- Andelić, S. (2010). *Prilog objektivnom vrednovanju rezultata rada studenata primenom računarskog adaptivnog testiranja*. Doktorska disertacija. Fakultet za industrijski menadžment, Union Univerzitet, Beograd, Srbija.
- Baker, F. B., & Kim, S. H. (2017). *The basics of item response theory using R* (pp. 17-34). New York: Springer.
- Beheshti, B., & Desmarais, M. (2012). Improving matrix factorization techniques of student test data with partial order constraints. In *International Conference on User Modeling, Adaptation, and Personalization* (pp. 346-350). Springer, Berlin, Heidelberg.
- Desmarais, M., Villarreal, A., & Gagnon, M. (2008, June). Adaptive test design with a naive bayes framework. In *Educational Data Mining 2008*.
- Hussain, S., Dahan, N. A., Ba-Alwib, F. M., & Ribata, N. (2018). Educational data mining and analysis of students' academic performance using WEKA. *Indonesian Journal of Electrical Engineering and Computer Science*, 9(2), 447-459.
- Kehoe, J. (1994). Writing multiple-choice test items. *Practical Assessment, Research, and Evaluation*, 4(1), 9.
- Nedelsky, L. (1954). Absolute grading standards for objective tests. *Educational and psychological measurement*, 14(1), 3-19.
- Scutari, M., & Denis, J. B. (2021). *Bayesian networks: with examples in R*. Chapman and Hall/CRC.
- Stefan, A. M., Schönbrodt, F. D., Evans, N. J., & Wagenmakers, E. J. (2022). Efficiency in sequential testing: Comparing the sequential probability ratio test and the sequential Bayes factor test. *Behavior Research Methods*, 1-18.
- Sumitha, R., Vinothkumar, E. S., & Scholar, P. (2016). Prediction of students outcome using data mining techniques. *International Journal of Scientific Engineering and Applied Science (IJSEAS)*, 2(6), 8.

## **KNOWLEDGE ASSESSMENT THROUGH THE APPLICATION OF COMPUTER ADAPTIVE TESTING IN RESPONSE TO MODERN TENDENCIES IN EDUCATION**

Svetlana Anđelić<sup>1,2</sup>, Nikola Dragović<sup>2</sup>, Goran Radić<sup>2</sup>, Valentin Kuleto<sup>2</sup>, Velimir Dedić<sup>1</sup>

<sup>1</sup>University PIM Banja Luka, Faculty of Computer Science, despota Stefana Lazrevića bb,  
78000 Banja Luka, Bosnia and Herzegovina

<sup>2</sup>Information Technology School ITS- Belgrade, Savski nasip 7, 11000 Belgrade, Serbia,  
[svetlana.andjelic@its.edu.rs](mailto:svetlana.andjelic@its.edu.rs)

### **ABSTRACT**

Education, at all levels, has undergone certain changes due to the Covid-19 pandemic. In the time of the pandemic, classes were held exclusively through online platforms or with a possible combination of traditional, frontal classes. Some institutions also conducted partial knowledge tests through online platforms with the possible inclusion of a video call meeting. The changes adopted at that time have largely remained even after the end of the pandemic. Online tests for self-assessment and assessment of knowledge are mainly based on a random question and answer generator. Such tests are not adapted to the current knowledge of the respondents themselves and thus negatively affect their motivation. Computer adaptive testing enables the test, ie the generated questions to be asked depending on the respondent's answer to the previously asked question (simulation of oral examination). This way of assessing knowledge has two positive effects, it gives the impression that the test has been created "just for him", while on the other hand it increases the effectiveness of testing because it reduces the number of questions to be answered to objectively assess the knowledge. This paper will describe the concept of an original, authorial CAT test model. The results of tests in the higher and secondary education system conducted using the above CAT model will also be presented.

**Keywords:** education, online test, CAT model, students, pupils.