

Mašinsko učenje u obrazovanju

Borivoje Milošević, Danica Milošević

Departman za SRT/KOT/GI

Visoka tehnička škola strukovnih studija

Niš, Srbija

borivojemilosevic@yahoo.com, danicamil@yahoo.com

Slobodan Obradović

Matematički institut

SANU

Beograd, Srbija

slobo.obradovic@gmail.com

Sadržaj—Mašinsko učenje u zadnjih nekoliko godina zauzima vrlo značajno mesto u svetu informacionih tehnologija, naročito kada su u pitanju analize velikog broja podataka (Data Market, Data Mining, Data Warehouse) i odgovori na upite iz ogromnih baza podataka: Google, Yahoo search, E-mail, itd. Međutim, mašinsko učenje može, a za to naravno u današnje vreme postoji i potreba, da se koristi i u obrazovanju. Široka je lepeza primene mašinskog učenja u obrazovanju, a u današnje vreme postoji i čitav niz aplikacija koje mogu u ovom slučaju pomoći korisniku. Koje su sve prednosti i koje mane učenja pomoću mašina, programiranih udžbenika i koji algoritmi i koje metode učenja postoje u svetu informacionih tehnologija? Kako se računar može koristiti u školstvu? Kako izgleda jedna uopštena blok šema algoritma za mašinsko učenje? Ostvareni rezultati i budućnost mašine - računara, kao predavača i ispitivača. To su osnovna pitanja na koja ćemo pokušati da odgovorimo u ovom radu.

Ključne reči - data mining; spam; algoritmi mašinskog učenja

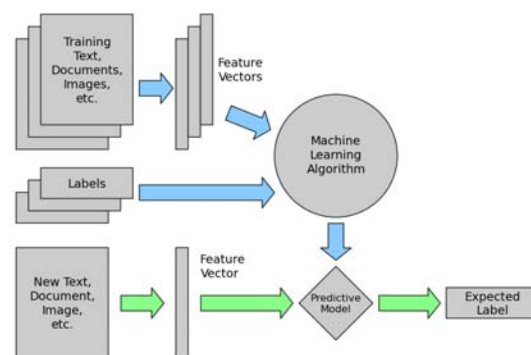
I. UVOD

Postoji čitav niz različitih problema koje je potrebno rešavati a koji i vrhunskim programerima zadaju skoro nerešive zadatke. Naprimera, kada je potrebno prepoznati otisak palca, prepoznati govor ili prepoznati lice osobe kao i neku od njegovih komponenti, sistem treba da sve to klasifikuje u jednu od mnogih kategorija i odluči koja je ta nepoznata osoba. Problem može biti i kod automatskog prevođenja dokumenata. Mogli bi za cilj potpunog razumevanje teksta pre prevođenje koristiti skup pravila izrađen od strane lingvista. Međutim, ovo je prilično naporan zadatak, posebno imajući u vidu da takav tekst nije uvek gramatički ispravan, niti je takvo razumevanje dokumenta moguće.

Umesto toga, mogli bi jednostavno koristiti primere prevedenih dokumenata, kao što je postupak kanadskog parlamenta ili drugih višezječnih subjekata (United Nations, Evropska unija, Švajcarska) da nauče mašinu kako prevesti između dva jezika. Drugim rečima, mogli bi koristiti te primere prevodilaca i na taj način naučiti mašinu kako da prevodi između njih, Sl 1. Ovaj pristup mašinskog učenja pokazao se prilično efikasan.

Mašinsko učenje se koristi kod pretraživanja Web stranica, kao spam filter kod e-mail sistema, kod sistema preporuke, analize tržišta, dizajniranju lekova, u školstvu itd. Pri tome se uglavnom koristi jedan od tipova mašinskog učenja - klasifikacija. To je sistem koji na svom ulazu ima vektor diskretnih ili kontinualnih vrednosti a za izlazni podatak samo jednu diskretnu vrednost, klasu. Tako naprimer spam filter klasifikuje E-mail poruke u "spam" ili "not spam," jer pošto je

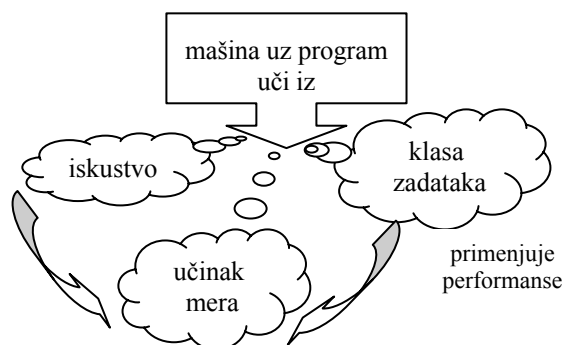
na ulazu Boolean vektor $x = (x_1, \dots, x_j, \dots, x_d)$, gde je $x_j = 1$ ako se reč u rečniku pojavljuje u E-mail sadržaju i $x_j = 0$ u drugom slučaju. Učenje je sadržano u setu primera (x_i, y_i) , gde je $x_i = (x_{i,1}, \dots, x_{i,d})$ ulazni podatak a y_i predstavlja korespondirajući klasifikovani izlaz. Drugačiji pristup rešavanju ovog problema bio bi vrlo složen i dugotrajan. Zato se može reći da se mašinsko učenje fokusira na predviđanju, na temelju poznatih svojstava naučenih iz skupa primera.



Slika 1. Proces primene mašinskog učenja u prevođenju teksta

Mašinsko učenje pripada oblasti veštačke inteligencije i odnosi se na izgradnju i proučavanje sistema koji mogu učiti i donositi logičke zaključke iz mase podataka, Sl. 1.

Kompjuterski program uči iz iskustva, u odnosu na neke klase zadataka i učinak mera, pri čemu realizovanje klase zadataka, poboljšano iskustvenim sadržajima na osnovu predviđenih mera, daje odgovarajuće rezultate, Sl. 2.



Slika 2. Proces mašinskog učenja

Ključna zamisao u ovom slučaju je, da ako želimo izbeći pisanje zametnih programskih linija za rešavanje postavljenih

zadataka i izmišljanja tople vode za svaku novu aplikaciju, primenom mašinskog učenja možemo smanjiti opseg različitih problema na set prilično uskih prototipova, koji će tada biti vrlo pogodni za razmatranje. To je osnovni zadatak i u organizovanju mašinskog učenja u školama.

Artur Samuel (1959) dao je definiciju mašinskog učenja: "polje proučavanja koje daje mašini - računaru mogućnost učenja ali bez eksplicitnog programiranja".

Postoji nekoliko vrsta mašinskog učenja [1]:

• *Induktivno učenje (Supervised learning)*: Podaci za obučavanje sadrže željeni izlaz

• *Samoobučavanje (Unsupervised learning)*: Podaci za obučavanje ne sadrže željeni izlaz

• *Semi-induktivno učenje*: Pored označenih primera, dominiraju neoznačeni primeri

• *Obučavanje sa pojačavanjem (Reinforcement learning)*: Nagrađivanje sekvence akcija

U literaturi nalazimo mnoštvo algoritama mašinskog učenja (Najbliži sused (Nearest neighbor), Grafovi/stabla odlučivanja, Neuronske mreže) ali su svi uglavnom bazirani na tri principa:

Reprezentacija. Klasifikator mora biti reprezentovan nekim od formalnih kompjuterskih jezika. Pri tome je izbor linearne reprezentacije ekvivalentan izboru seta klasifikatora na osnovu kojih postoji mogućnost učenja. Ovaj set se zove hypothesis learning space. Ako klasifikator nije u ovom području, ne može se učiti.

Procenjivanje. Funkcija procenjivanja (objective function) je potrebna da bi se razlikovali dobri od loših klasifikatora.

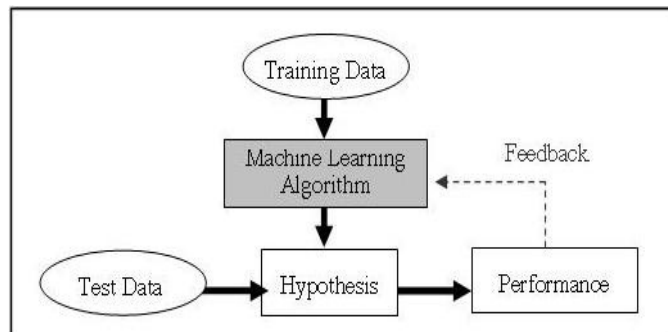
Optimizacija. Potreban je metod pretraživanja klasifikatora u okviru najvećih vrednosti, pa tako izbor optimizacije tehnike predstavlja ključ za produktivno učenje, a takođe pomaže u determinisanju klasifikatora ako funkcija procenjivanja ima više od jedne optimalne vrednosti, tabela I.

TABELA I. KONPONENTE ALGORITAMA MAŠINSKOG UČENJA

Representation	Evaluation	Optimization
Instances	Accuracy/Error rate	Combinatorial optimization
K-nearest neighbor	Precision and recall	Greedy search
Support vector machines	Squared error	Beam search
Hyperplanes	Likelihood	Branch-and-bound
Naive Bayes	Posterior probability	Continuous optimization
Logistic regression	Information gain	Unconstrained
Decision trees	K-L divergence	Gradient descent
Sets of rules	Cost/Utility	Conjugate gradient
Propositional rules	Margin	Quasi-Newton methods
Logic programs		Constrained
Neural networks		Linear programming
Graphical models		Quadratic programming
Bayesian networks		
Conditional random fields		

Algoritmi mašinskog učenja imaju za zadatak da se obave važni poslovi njihovom generalizacijom na osnovu primera i iskustva. Ovo je često moguće i manje košta jer programiranje

ovakvih zadataka zahteva mnogo programerskog truda i znanja, ali u ovom slučaju sa nesigurnim ishodima. Uz to, zbog prisutnosti ogromnog broja podataka koje treba analizirati, javljaju se mnogo ambiciozniji problema za rešavanje. Kao izvor, mašinsko učenje koristi računarske, statističke i filozofske nauke a razvoj i projektovanje uspešnih aplikacija mašinskog učenja zahtevaju solidnu količinu sveukupnih znanja koja se mogu teško naći u literaturi. Proces mašinskog učenja dat je na Sl. 3.



Slika 3. Proces mašinskog učenja

II. MAŠINSKO UČENJE U OBRAZOVANJU

Teorija programirane nastave, čiju su osnovu izgradili Presi, Skinner i Krauder [2] i koja je nastala kao neumitna etapa u razvoju pedagogije, psihologije i informatike, po mnogim autorima radova iz ove oblasti današnjice predstavlja jedini put za što kvalitetnije i kvantitetnije usvajanje mnoštva informacija koje se javljaju u samom procesu učenja.

Osnovni princip mašinskog učenja i programirane nastave sastoji se u tome da nastavnik, u uslovima masovne nastave, prenese deo svojih funkcija na specijalne uređaje za učenje, koji u tom slučaju omogućuju ispunjenje svih onih bitnih uslova koji su inače mogući samo u individualnom radu. A to su:

1. istovremeno prihvatanje informacija od svakog učenika pojedinačno, njihovo analiziranje i shvatanje;
2. istovremeno podučavanje svakog učenika, koristeći pri tome metode koje su optimalne za svakog pojedinog učenika;
3. istovremeno obavljanje nastave tempom koji odgovara svakom učeniku pojedinačno;
4. istovremeno stvaranje rešenja o optimalnim načinima obučavanja raznih učenika;
5. istovremeno ostvarivanje neposrednog proveravanja i obaveštavanja učenika o tačnosti svakog njegovog odgovora, stvarajući pri tome pozitivnu motivaciju svakog učenika na ono šta uči;
6. istovremeno obnavljanje onih informacija posle kojih je svaki pojedinačni učenik davao netačne odgovore, onoliko puta koliko je to svakom učeniku potrebno da bi shvatio dati problem (informaciju) i odgovorio tačno na postavljena pitanja;
7. istovremeno održavanje što veće pažnje i aktivnosti svakog učenika u toku rada;

8. izbegavanje svih subjektivnih faktora nastavnika, koji su uticali na ispitivanje i ocenjivanje učenikovog znanja, odnosno na određivanje nivoa savladanosti programiranog gradiva.

Za samo izvođenje programirane nastave [2] potrebno je rešiti tri problema. Prvi se sastoji u sastavljanju programiranog udžbenika, imajući u vidu samu nastavno-motodičku osetljivost gradiva za učenje. Koristeći tehniku linearnog ili razgranatog programiranja udžbenika možemo programirati neki od običnih udžbenika na taj način što ćemo celokupno gradivo raspoređati na niz malih logičkih celina (jediničnih informacija) [3]. Postojanje između njih naziva se "korak mašinskog učenja" i u zavisnosti od modela može biti veći ili manji. Posle svake jedinične informacije učeniku se prezentuje pitanje i daje mogućnost da na njega odgovori, izabirajući jedan od ponudjenih odgovorai ili izgrađujući sopstven odgovor. Posle odgovaranja na postavljeno pitanje učeniku se prezentuje da li je odgovorio tačno, i u zavisnosti od toga daje mu se nova jedinična informacija iz logičkog niza ili ukazuje na grešku i objašnjava njena priroda, vraćajući ga na ponovo proučavanje informacije posle koje je dao pogrešan odgovor.

Medjutim, polazeći od mogućnosti da je običan udžbenik u pogledu izlaganja težak i ne odveć razumljiv, gradivo za učenje neće postati lakše i razumljivije time što je programirano, pa je potrebno preraditi postojeći udžbenik izradom jednog savršenijeg modela, sada pogodnijeg za programiranje.

Programirani udžbenik možemo smatrati već poluautomatskom "mašinom" za učenje, jer on na osnovu provere usvojenosti prethodne informacije daje učeniku radne informacije po kojima treba on dalje da se vladala.

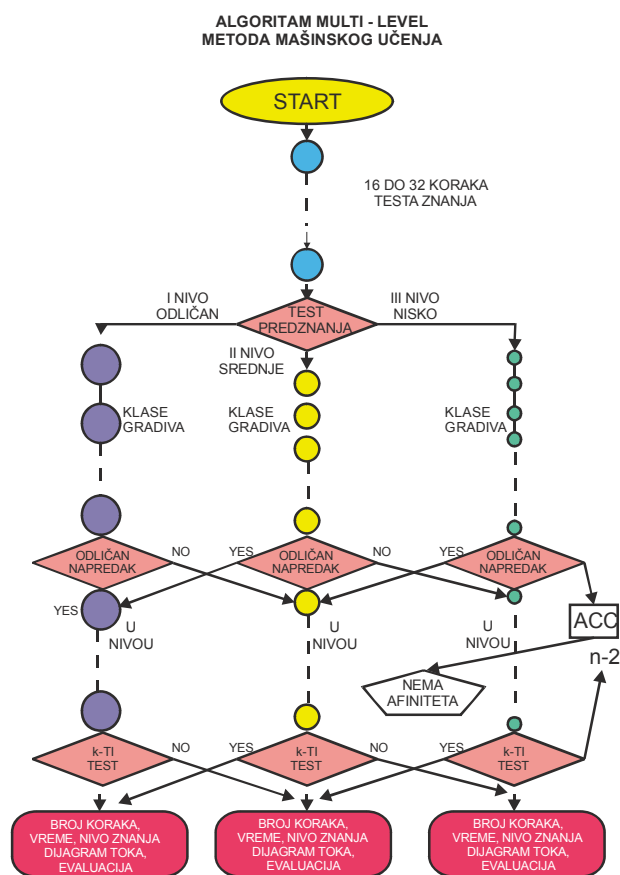
Drugi problem je sastavljanje programa i postizanje njegove algoritamsko - logičke korektnosti. Ovaj problem je znatno lakše rešiti jer on zavisi samo od umešnosti programera i njegovog poznavanja tehnike programiranja. Obično se piše u nekom od interaktivnih jezika i predviđa rad sa displejima, tastaturom i drugim multimedijalnim uređajima, kao ulazno-izlaznim jedinicama. Izlazni podaci, koji se daju posle svakog obrađenog poglavlja i na kraju programa, moraju pored izveštaja o učinjenim greškama, broju prolaza kroz programirano gradivo i vremena utrošenog za učenje (radi utvrđivanja nivoa savladanosti programiranog gradiva) sadržati i dijagnostičke podatke koji bi služili za utvrđivanje nedostataka programa analizom učenikovih odgovora. Tako bi se revizijama samog programa mogla odstraniti "uska grla" i "široka mesta" u savladjivanju programiranog materijala i dobiti optimalni programi koji bi prema našem iskustvu omogućili velikom broju učenika da usvoje čitavo programirano gradivo sa maksimalnim učinkom i za minimum vremena, Sl. 4.

Treći problem je fizičko organizovanje same učionice za programirano učenje.

Studenti i drugi korisnici imaju brojne prednosti korišćenjem sistema za mašinsko učenje. Mogu pokrenuti on-line kurseve, prisustvovati on-line ispitima, dobijati povratne informacije od profesora i instruktora i slati svoje projekte i seminarske radove njima na ocenjivanje. Sistem se može lako prilagoditi potrebama više korisnika ili dodatnih usluga – ili se

može smanjiti aktivnosti kad potražnja za uslugama sezonski opadne [4].

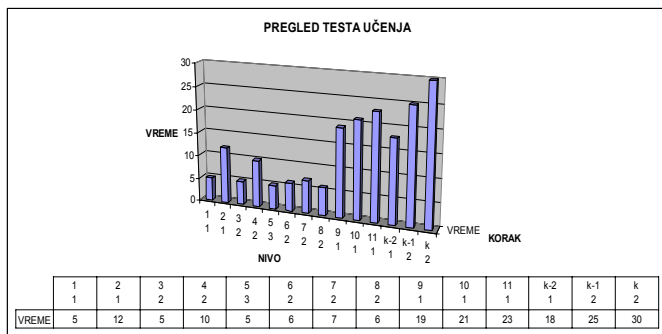
Prednosti za predavače: Ogromne prednosti za predavače omogućuje primena ovakvog sistema. Njima se nudi priprema on-line studentskih testova, bavljenje i kreiranje boljih resursa za studente kroz upravljanje sadržajima, pristup testovima i njihovoj izradi, pristup sadržajima kućnog rada studenata, projektima koje su studenti uradili, slanje povratnih informacija i komunikacija sa studentima kroz on-line forume.



Slika 4. Algoritam mašinskog učenja

Koristeći sve ove podatke nastavnik može lako da stekne uvid u to kako je učenik savladao programirani materijal, gde je napravio greške u odgovaranju, a upoređujući više dijagrama sa rezultatima testa za svakog studenta pojedinačno može uočiti mesto gde se sam program za učenje mora menjati, gde treba povećati ili smanjiti težinu informacija i dužinu koraka, čineći na taj način da programirani materijal i program za upravljanje daju optimalne rezultate, Sl. 5.

U ovom slučaju upotrebljavaju se mašine i programi koji se prilagođavaju okruženju koje se menja vremenom (dinamički sistemi) [5]. Oni su čest i potreban zahtev u mnogim oblastima pa i u procesu mašinskog učenja, jer je za mnoge zadatke cilj naći rešenje koje se fleksibilno prilagođava, umesto da se koriste metode veštačke inteligencije kojima je potreban reinženjering nakon svake suštinske promene.



Slika 5. Test mašinskog učenja

Metodi mašinskog [6] učenja se vrlo uspešno koriste u rešavanju brojnih zadataka razvoja informacionih sistema i to u prikupljanju korisničkih zahteva (izvlačenje znanja, izrada prototipova), estimaciji mogućnosti ponovnog korišćenja softverskih modula (aplikacije i generatori), testiranju i validaciji sistema, održavanju (razumevanje softvera) i upravljanju projektima razvoja (procena cene i trajanja projekta razvoja, obima sistema i predviđanje ili procena greški u sistemu), [7] tabela II.

TABELA II. METODE MAŠINSKOG UČENJA

METODE	neuronske mreže	na osnovu instanci	stablo odlučivanja	genetički algoritmi	logička pravila	reformulacija	deduktivni	noseći vektori
Predviđanje i estimacija	*	*	*	*	*			
Otkrivanje svojstava i modela	*				*	*		*
Transformacija	*	*		*				
Generisanje i sinteza		*	*	*	*	*	*	
Ponovno korišćenje softvera	*	*	*	*				*
Prikupljanje			*		*	*		
Upravljanje		*						

Posmatrajući čitav niz novijih algoritama za mašinsko učenje u radu ćemo se zadržati samo na onim najboljim [15]. To se sa sigurnošću može utvrditi, upoređivanjem njihovih međusobnih odnosa. U tabeli III dat njihov međusobni odnos. Može se zaključiti da je Naive Bayes algoritam pobeđuje u tri koraka, u 6 nema značajnih statističkih razlika i nema ni jednu značajnu statističku grešku.

TABELA III. POREĐENJE ALGORITAMA

Algorithm	3-NN	RIPPER	C4.5	WINNOWER
Naive Bayes	0/6/3	0/9/0	0/5/4	0/0/9
3-NN		2/7/0	2/4/3	0/1/8
RIPPER			1/5/3	0/1/8
C4.5				0/4/5

Sa ove tačke gledišta, Naive Bayes algoritam predstavlja onaj koji najviše odgovara učenju, pa se tako može koristiti kao postupak pri projektovanju softvera za mašinsko učenje. Prototipna verzija ovog softvera urađena je na Department of Mathematics, University of Patras, Greece, a sa Sl. 6. se može videti njegova primena u toku jedne akademske godine [15].

Slika 6. Izgled prozora softverskog paketa zasnovanog na Naive Bayes algoritmu

ZAKLJUČAK

Mašinsko učenje u obrazovanju, po mišljenju mnogih stručnjaka iz te oblasti, studentima može pomoći više nego većina drugih oblasti računarskih nauka. Njima je danas već sasvim jasno, da kompjuterske nauke i mašinsko učenje predstavljaju jedno vrlo izazovno i uzbudljivo naučno polje. Konkretno, napredak u rešavanju "velikih problema" (razumevanje ćelije, lečenje raka, pronalaženje crne materije između galaksija) strogo zavisi danas od najmodernije tehnologije računarskih nauka i mašinskog učenja. I tako, kako vreme prolazi, brojni izazovi se javljaju u raznim oblastima njegove primene. Identifikovanje, pripremanje i testiranje ovih zadataka biće jedna od dugotrajnih i vrlo delikatnih faza u razvoju mašinskog učenja u školstvu.

ZAHVALNICA

Ovaj rad je realizovan u okviru projekta III 44006 Ministarstva obrazovanja i nauke Republike Srbije.

LITERATURA

- [1] Pedro Domingos, A Few Useful Things to Know about Machine Learning, Department of Computer Science and Engineering, University of Washington, Seattle, WA 98195-2350, U.S.A
- [2] Danica Milošević, Electronic Learning of Business English, 23rd British and American Studies Conference, The West University of Timisoara, English Department, 16th-18th May, 2013, The West University of Timisoara, Romania
- [3] Vladislav Miškovic, Metodi mašinskog učenja, i DSS sistemi za podršku odlučivanju, Fakultet za informatiku i računarstvo, Beograd

- [4] Alex Smola and S.V.N. Vishwanathan, Introduction to Machine Learning, Yahoo! Labs, Santa Clara
- [5] Rainer Marrone, Ralf Möller, Foundations of Machine Learning and Data Mining,
- [6] Ricard Gavalda, Machine Learning in Secondary Education, Universitat Politècnica de Catalunya (UPC), Barcelona (Spain)
- [7] Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar, Foundations of Machine Learning, (2012), The MIT Press.
- [8] Arroyo, I. Woolf, B. Inferring learning and attitudes from a Bayesian Network of log file data Proceedings of the 12th International Conference on Artificial Intelligence in Education. C.K. Looie, G. McCalla, B. Bredeweg, and J. Breuker, editors. Amsterdam: IOS Press, 2005. pp 33-40.
- [9] Arroyo, I. Martin, K. AgentX: Using Reinforcement Learning to Improve the Effectiveness of Intelligent Tutoring Systems. 564-572. James C. Lester, Rosa Maria Vicari, Fábio Paraguaçu (Eds.): Intelligent Tutoring Systems, 7th International Conference, ITS 2004, Maceió, Alagoas, Brazil, August 30 - September 3, 2004, Proceedings. Lecture Notes in Computer Science 3220 Springer 2004.
- [10] Ferguson, K., Arroyo, I., Mahadevan, S, Woolf, B., Barto, A. Improving Intelligent Tutoring Systems: Using Expectation Maximization To Learn Student Skill Levels. Proceedings of the Intelligent Tutoring Systems, Taiwan, June, 2006.
- [11] Johns, J. Mahadevan, S. Woolf, B.P. Estimating Student Proficiency Using an Item Response Theory Model
- [12] Arroyo, I. Barto, A. Ferguson, K. Mahadevan, S. Woolf, B.P. Improving Intelligent Tutoring Systems: Using Expectation Maximization To Learn Student Skill Levels. Proceedings of the Intelligent Tutoring Systems, Taiwan, June, 2006.
- [13] Arroyo, I Barto, Fisher, D. A. Johns, J. Jonsson, A. Mahadevan, S. Mehranian, H. Woolf, B.P. Evaluating the Feasibility of Learning Student Models from Data
- [14] Beck, J.E. Beal, C.R. Woolf, B.P. ADVISOR: A machine learning architecture for intelligent tutor construction
- [15] S. B. Kotsiantis, C. J. Pierrakeas, I. D. Zaharakis, P. E. Pintelas, EFFICIENCY OF MACHINE LEARNING TECHNIQUES IN PREDICTING STUDENTS' PERFORMANCE IN DISTANCE LEARNING SYSTEMS, Educational Software Development Laboratory Department of Mathematics, University of Patras, Greece
- [16] Mitchell, T. (1997), Machine Learning, McGraw Hill.
- [17] Murthy, S. (1998), "Automatic Construction of Decision Trees from Data: A Multi-Disciplinary Survey", Data Mining and Knowledge Discovery, Vol. 2, pp. 345-389.
- [18] Quinlan, J. R. (1993), C4.5: Programs for machine learning, Morgan Kaufmann, San Francisco
- [19] Salzberg, S. (1997), "On Comparing Classifiers: Pitfalls to Avoid and a Recommended Approach", Data Mining and Knowledge Discovery, Vol. 1, pp. 317-328.
- [20] Wettschereck, D., Aha, D. and Mohri T. (1997), "A Review and Empirical Evaluation of Feature Weighting Methods for a Class of Lazy Learning Algorithms", Artificial Intelligence Review, Vol. 10, pp. 1-37.
- [21] Witten, I. and Frank E. (2000), Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufmann, San Mateo.
- [22] Xenos, M., Pierrakeas C. and Pintelas P. (2002), "A survey on student dropout rates and dropout causes concerning the students in the course of informatics of the Hellenic Open University", Computers & Education, Vol. 39, pp. 361-377.

ABSTRACT

Machine learning has been occupying a very important place in the world of information technology in recent years, especially when it comes to analyzing large amounts of data (Data Market, Data Mining, Data Warehouse) and respond to queries from huge databases such as: Google, Yahoo Search, E - mail, etc. However, machine learning can be, and of course nowadays it is used in education due to the great need for it. Nowadays, there is a wide variety of applications in educational machine learning, and there are many applications that can be helpful to users. What are the advantages and disadvantages of using machine learning, programmed textbooks, the algorithms and methods for learning that exist in the world of Information Technology? Can the computer be used in education? How does the algorithm of a general block diagram for machine learning look like? What are the actual results and what is the future of machines - computers as teachers and examiners. These are the basic questions that we are going to give answers to in our paper.

EDUCATIONAL MACHINE LEARNING

Borivoje Milošević, Slobodan Obradović, Danica Milošević