

ELENA ȘUȘNEA

**UTILIZAREA
TEHNICILOR DATA MINING
ÎNTR-UN SISTEM
EDUCAȚIONAL
DE TIP E-LEARNING**



Copyright © 2012, **Editura Pro Universitaria**

Toate drepturile asupra prezentei ediții aparțin
Editurii Pro Universitaria

Nicio parte din acest volum nu poate fi copiată fără acordul scris al
Editurii Pro Universitaria

Descrierea CIP a Bibliotecii Naționale a României

ȘUȘNEA, ELENA

**Utilizarea tehnicilor data mining într-un sistem
educațional de tip e-Learning / Elena Șușnea. - București : Pro
Universitaria, 2012**

Bibliogr.

ISBN 978-606-647-543-3

004:371

004.43 Data Mining

CUVÂNT ÎNAINTE

În prezent, educația reprezintă un factor hotărâtor în dezvoltarea pe termen lung a unei țări, implicând bunăstarea și progresul acesteia. Națiunile care aleg sistemele de instruire ce utilizează noile tehnologii educaționale vor fi națiunile cele mai puternice din punct de vedere economic în secolul XXI.

Digitizarea, interactivitatea și virtualitatea oferă noi oportunități și extind limitele privind modul de organizare și desfășurare a proceselor educaționale la toate nivelurile, fiind factori importanți ai schimbărilor ce au loc în domeniul educațional militar.

Contextul apariției acestei cărți este propice, atât datorită creșterii interesului studenților manifestat la nivel global, pentru cursurile desfășurate online, cât și exigențelor NATO privind utilizarea unui sistem avansat de instrumente și tehnologii de instruire pentru profesionalizarea personalului armatei.

Dinamica înregistrată de noile tehnologii va avea în continuare un impact profund asupra premiselor de bază ale învățământului militar și va da naștere la căutarea de teorii și metodologii educaționale noi, care să satisfacă cerințele privind predarea, învățarea și cercetarea din domeniul academic militar.

Platformele educaționale LMS (Learning Management System, sistem de management al învățării), acele aplicații informatice complexe cu facilități precum organizarea și furnizarea de informații, sprijinirea proceselor de învățare, supervizarea și evaluarea rezultatelor studenților, permit stocarea unui volum mare de date. Exploatarea eficientă a acestor date contribuie semnificativ la dezvoltarea unor instrumente care să faciliteze realizarea unui învățământ centrat pe student.

Scopul acestei cărți este înțelegerea proceselor desfășurate într-un mediu virtual de învățare prin analiza datelor furnizate de un sistem de management al învățării și elaborarea unor predicții privind aceste procese. În acest sens, autoarea cărții prezintă într-o formă integratoare unele metode de analiză preluate din statistică și inteligență artificială, reușind o îmbinare perfectă între cunoșterea științifică, experiența e-learning și practica specifică domeniului educațional militar.

Cultura și talentul didactic al autoarei sunt evidente. Utilizarea tehnicilor din domeniul data mining, un domeniu științific transdisciplinar relativ nou, este ilustrată în susținerea unor activități de tip e-learning folosind pentru exemplificarea aspectelor metodologice o aplicație cu caracter didactic.

Materialul este bine organizat, prezentarea este gradată, expunerea este foarte elegantă și în același timp, clară și inteligibilă. Lucrarea se concentrează pe crearea unei baze teoretice prin prezentarea tehnicilor data mining utilizate în mediul de învățare virtual și pe elaborarea unor modele care să genereze pe termen lung cunoștințe utile studenților, profesorilor sau tutorilor, managerilor educaționali și dezvoltatorilor de produse software educaționale, continuând cu valorificarea acestor modele în procesele din cadrul și din afara sistemelor de management al învățării.

Data mining este un proces laborios în care sunt utilizate metodele de descriere și predicție pentru a elabora modele care ulterior sunt testate pentru a elabora informațiile necesare proceselor decizionale.

Analiza datelor în vederea obținerii de informații recurge la diverse tehnici, printre cele mai folosite aflându-se: clustering, reguli de asociere, arbori de decizie, rețele neuronale artificiale. Deși unele tehnici data mining datează de mai mulți ani, cum sunt cele preluate din statistică, totuși algoritmi folosiți au cunoscut un proces de evoluție continuă, care a permis înlăturarea unora dintre limitele sau deficiențele inițiale.

Perspectiva adoptată în această lucrare de autoarea Elena Șuşnea este aceea că tehnologiile data mining vor contribui semnificativ la îmbunătățirea capacităților de învățare ale studenților. Prin înglobarea tehnicilor din inteligența artificială și statistică în sistemele de tip e-learning se vor crea premisele dezvoltării unor sisteme inteligente de instruire.

O asistență inteligentă și personalizată pe care un profesor sau tutore o poate oferi într-o anumită situație educațională în clasă (tradițională sau virtuală) nu este ușor de realizat. Prin utilizarea tehnologiilor data mining în sistemele de management al învățării se adaugă funcționalitate în sensul asistării și adaptării inteligente a sistemului la situația creată.

Instrumentele ce pot fi dezvoltate pe baza modelelor rezultate în urma folosirii tehnicilor data mining nu elimină nevoia de a avea profesori sau tutori comuni însă rolul acestora se modifică. Aceștia trebuie să supravegheze și să evalueze crearea materialelor necesare în instruirea on-line, să asiste studenții și să

monitorizeze progresul înregistrat de aceștia, să revizuiască materialele didactice în funcție de feedback-ul primit de la studenți etc. În plus, interacțiunea umană este inestimabilă în multe situații de instruire on-line.

În viitor, odată cu utilizarea tehnicilor data mining în educație, o mare parte din funcțiile profesorului sau al tutorelui vor fi preluate de agenți software care pot îndeplini această funcție cu scopul de a ajuta studentul. Astfel, e-learning va deveni un mediu educațional ce va permite instruirea rapidă a unui număr din ce în ce mai mare de studenți într-un mod mult mai eficient.

Valoarea lucrării este în primul rând didactică atât prin prisma modelului de sistematizare și prezentare a metodelor de minerit ilustrate cu exemple didactice, cât și prin cea a încercărilor de a contribui la îmbunătățirea predării cu ajutorul mijloacelor TIC (tehnologia informației și comunicațiilor) prin furnizarea de feedback profesorului/instructorului în faza de examinare – evaluare a cursanților.

Concluziile trase prin experimentarea pe o bază de date reală sunt foarte valoroase, ele contribuind la înțelegerea unor efecte generate de o parte dintre activitățile susținute prin intermediul platformelor informatice de tip e-learning.

Acumularea și crearea de cunoștințe în cadrul proceselor educaționale implică antrenarea unor resurse variate (resurse umane precum profesori sau tutori, resurse financiare, echipamente hardware, produse software educaționale, timp etc.) în scopul realizării obiectivelor didactice. Însă nu este suficient să dispui de cele mai moderne tehnologii pentru a desfășura un proces educațional optim, trebuie să accesezi materialele educaționale, să-ți dorești să studiezi și să ai competențe metacognitive necesare să o faci.

Teodor FRUNZETI

INTRODUCERE

Domeniul educațional a cunoscut în ultimii ani transformări profunde în special datorită influenței noilor tehnologii educaționale. Schimbările importante produse de acestea în sistemul de învățământ ating un stadiu care permit universităților să desfășoare un proces educațional eficient, adaptat nevoilor studenților.

Odată cu creșterea popularității Internetului, instituțiile de învățământ superior au început să diversifice oferta de servicii educaționale pentru categoriile de studenți. Un studiu publicat în noiembrie 2008, efectuat de Allen E., Seaman J. ș.a. [1] de la consorțiul Sloan-C, în care se examinează starea învățământului universitar on-line desfășurat în S.U.A. [2], indică o creștere a interesului manifestat de studenți pentru această formă de învățământ. Astfel, comparativ cu datele publicate în anul 2004, raportul curent indică existența unui număr de două ori mai mare de studenți ce urmează cursuri on-line și o rată de creștere de 12.9 % a numărului de studenți pentru anul universitar 2007-2008 față de anul precedent.

Apariția și dezvoltarea noilor tehnologii a avut un impact puternic și asupra învățământului militar. Evoluția proceselor și fenomenelor militare este strâns corelată cu dezvoltarea tehnologică. Profesionalizarea personalului armatei presupune, printre altele, utilizarea unui sistem avansat de instrumente și tehnologii de instruire, inclusiv învățarea distribuită avansată la distanță.

Unul din principiile care stau la baza modernizării Armatei României, principiu prevăzut și în Strategia Militară a României [3], este promovarea tehnologiilor moderne. Utilizarea acestor tehnologii în procesul instruirii va permite operaționalizarea forțelor conform standardelor Alianței și participarea acestora la îndeplinirea cu succes a misiunilor pentru care au fost create.

Prin introducerea standardelor impuse de Învățământul Distribuit Avansat la Distanță (Advanced Distributed Learning, ADL¹) în armată, resursele educaționale devin accesibile oriunde și oricând [4]. Aceasta înseamnă că se pot

¹ Strategie elaborată în cadrul Departamentului Apărării al Statelor Unite care se ocupă cu standardizarea și modernizarea educației și formării profesionale prin utilizarea tehnologiilor informației și comunicațiilor .

asigura mobilitatea, operativitatea personalului și realizarea de economii substanțiale, caracteristici foarte importante pentru domeniul militar.

Prin folosirea sistemului e-learning, personalul militar va avea acces la un sistem educațional modern, de cea mai înaltă calitate, conceput pentru nevoile specifice din domeniu.

În prezent, sistemele e-learning permit achiziția și stocarea unui volum imens de date. Exploatate suficient, aceste date conduc la obținerea unor informații care să reflecte comportamentul sistemului analizat și, în consecință, a procesului educațional desfășurat în cadrul instituțiilor de învățământ.

O platformă e-learning completă trebuie să fie un sistem integrat de predare, învățare și gestiune a conținutului educațional bazat pe principii pedagogice moderne și să ofere suport pentru predare și învățare, testare și evaluare, administrarea conținutului, monitorizarea procesului de învățământ și concepția curriculară.

Universitatea Națională de Apărare “Carol I” folosește platforma ILIAS¹, din anul 2006, oferind un pachet de cursuri on-line cu suport SCORM² (Sharable Content Object Reference Model) care permite monitorizarea și parcurgerea conținuturilor educaționale conform nevoilor fiecărui student [5]. De asemenea, platforma e-learning oferă informații factorilor responsabili din domeniile decizional, control și planificare respectiv proiectare software, privind modul de desfășurare a procesului de învățământ [6].

Informațiile necesare pentru analiza procesului de învățământ pot fi extrase din datele existente în baza de date a platformei ILIAS, însă acestea sunt dificil de extras. Tehnicile din domeniul *Knowledge Discovery in Database* (KDD) și *Data Mining* (DM) ajută utilizatorii să descopere informații folositoare în baza de date. Deși din 1990 DM și KDD au fost subiecte de interes pentru comunitatea științifică, totuși terminologia încă variază și definițiile nu sunt întotdeauna clarificatoare.

¹ Sistem de management a învățării bazat pe Web, care permite gestionarea facilă a resurselor într-un sistem integrat, detalii la <http://www.ilias.de>.

² SCORM integrează un set de standarde tehnice, specificații, precum și orientări concepute pentru a satisface exigențele utilizatorilor privind accesibilitatea, interoperabilitatea, durabilitatea și reutilizarea conținutului și a sistemelor de management al învățării. Această colecție de standarde este definită de Advanced Distributed Learning.

Într-un sistem educațional centrat pe student, datele privitoare la activitățile desfășurate de studenți pe toată perioada derulării cursului la care sunt înscriși, prelucrate și interpretate corect, pot furniza informații importante personalului responsabil de buna desfășurare a activităților didactice (manageri educaționali, profesori, dezvoltatori software, specialiști din domeniul psiho-pedagogiei).

Platforma ILIAS furnizează în acest sens, două tipuri de date:

- resurse educaționale - sistemul pune la dispoziție o serie de resurse stocate în server, incluzând pagini Web, prin intermediul cărora sunt prezentate conținuturile educaționale, temele pentru acasă, testele parțiale și finale, link-uri către alte resurse educaționale disponibile în spațiul Web, simulări;
- date referitoare la activitățile utilizatorilor, cum ar fi: crearea, modificare, evaluarea sau folosirea resurselor educaționale, momentul logării / delogării.

Printre factorii care determină creșterea cantității de date stocate în sistemul ILIAS menționăm:

- existența unor cursuri în desfășurare, implică stocarea unor date, pentru fiecare student, privind accesarea suportului de curs, parcurgerea testelor de evaluare, înregistrarea răspunsurilor și a numărului de încercări efectuate pentru fiecare item etc.;
- apariția unor cursuri noi, presupune pe lângă stocarea datelor amintite anterior și crearea unor noi resurse educaționale în conformitate cu profilul absolventului;
- actualizarea conținuturilor educaționale, prin înlocuirea acelor părți considerate “depășite” cu unele care corespund unor rezultate științifice recente și au aplicabilitate mai mare;
- diversificarea formatului electronic de prezentarea a materialelor didactice, prin substituirea unor formate statice cu unele care asigură un grad mare de interactivitate și sunt mai sugestive.

Având în vedere cele menționate mai sus, ne propunem să rezolvăm un aspect important privind extragerea unor cunoștințe din seturile de date ale studenților care au desfășurat cursuri de pregătire on-line accesând resursele educaționale puse la dispoziție de platforma e-learning ILIAS. În procesul de

extragere a unor informații noi și utile pentru cunoașterea proceselor educaționale, am utilizat metodele și tehnicile data mining.

Organizare și conținut

Prima parte a cărții prezintă etapele pe care le-a traversat educația la distanță din punct de vedere a transmiterii informațiilor către studenți. Acest tip de educație s-a dezvoltat și rafinat continuu atingând punctul culminant odată cu asimilarea noilor tehnologii ale informației și comunicațiilor, conducând la paradigma „e-learning”. Creșterea popularității Internetului a determinat apariția învățământului bazat pe Web (Web-Based Learning) și odată cu el dezvoltarea și diversificarea sistemelor de management al învățării (Learning Management System). Sunt prezentate deasemenea categoriile de date furnizate de platforma e-learning ILIAS, factorii care determină creșterea volumului de date și modul în care pot fi analizate aceste date pentru extragerea informațiilor utile dezvoltării unui sistem de învățământ centrat pe student.

În **primul capitol**, intitulat ”Sisteme educaționale de tip e-learning”, sunt analizate trăsăturile comune ale platformelor educaționale de tip e-learning și factorii care sunt luați în considerare când se achiziționează o astfel de platformă, un factor important fiind numărul de utilizatori. Deși există peste 50 de sisteme open-source pentru managementul învățării, printre cele mai utilizate sunt Moodle și ILIAS.

Deoarece analiza făcută în cadrul acestei cărți are la bază date furnizate de sistemul ILIAS, au fost prezentate detaliat caracteristicile și facilitățile acestuia. Sistemul ILIAS are o arhitectură modulară, oferind un set de servicii la nivelul interfeței utilizator datorită existenței următoarelor subsisteme: subsistemul de evaluare, subsistemul de administrare a învățării, subsistemul de gestionare a bazelor de date, subsistemul de particularizare a interfețelor. De asemenea, sunt făcute câteva considerații privind implementarea unui modul software pentru analiza datelor furnizate de platformă.

Capitolul 2, denumit ”Data mining și procesul de descoperire a cunoștințelor”, are ca scop prezentarea domeniului științific data mining. Asaltați de volumul tot mai mare de date conținute de bazele de date, cercetătorii din domeniile bazelor de date, statisticii, inteligenței artificiale și recunoașterea formelor, au început să definească proceduri standard care să ghideze utilizatorii în extragerea informațiilor utile cu scopul cunoașterii sistemului descris de acele

date. În acest sens au fost elaborate mai multe modele, cel mai des utilizat fiind modelul CRISP-DM, propus de un mare consorțiu de companii europene (Integral Solution Ltd., NCR, DaimlerChrysler, OHARA) care constă în parcurgerea a șase faze. Extragerea informațiilor din bazele de date poate avea ca obiective: modelarea descriptivă a sistemului, analiza exploratorie a datelor, modelarea predictivă, descoperirea pattern-urilor și a regulilor, căutarea după conținut. În acest sens, sunt prezentate metodele data mining și principalele aplicații software care utilizează aceste metode.

Deși metodele și tehnicile din domeniul data mining sunt din ce în ce mai des utilizate, totuși terminologia încă mai variază iar procesul data mining este frecvent confundat cu procesul de descoperire a cunoștințelor din bazele de date (Knowledge Discovery in Database, KDD). În 1996, Fayyad ș.a. definește pentru prima dată conceptul KDD și elaborează un model de bază pentru extragerea cunoștințelor din bazele de date. În acest model, data mining este prezentat ca subproces al procesului KDD (alături de selecția, preprocesarea, transformarea datelor respectiv interpretarea-evaluarea informațiilor).

Capitolul 3, intitulat ”Metode de extragere și selecție a caracteristicilor”, prezintă metode de extragere și selecție a caracteristicilor obiectelor stocate în baza de date. Cele două operații sunt importante deoarece permit reducerea complexității datelor și implicit eliminarea zgomotului, reducerea memoriei alocate și a timpului de procesare a acestora. În funcție de numărul de caracteristici de intrare, tehnicile de extragere a caracteristicilor poate aduce o îmbunătățire substanțială a algoritmului de calcul din punct de vedere al complexității acestuia.

La finalul acestui capitol, este realizată o analiză exploratorie a datelor conținute de baza de date furnizată de platforma educațională ILIAS, ca pas premergător în identificarea claselor prin utilizarea metodelor de clustering prezentate în capitolul următor.

Capitolul 4 cu titlul ”Metode și tehnici data mining utilizate în modelarea descriptivă și descoperirea regulilor de asociere” prezintă tehnicile de clustering utilizate în instrumentele de tip data mining cu exemplificări pe baza de date. S-a evidențiat faptul că, deoarece acestea nu sunt supervizate, gradul lor de clasificare a datelor rămâne puternic dependent de distribuția caracteristicilor din punctul de vedere al distanțelor interclase și intraclasă. Sunt aplicate tehnici precum clustering și algoritmul k-means.