



Universitatea
Ștefan cel Mare
Suceava

FACULTATEA DE INGINERIE
ELECTRICĂ ȘI ȘTIINȚA
CALCULATOARELOR

REZUMAT

CONTRIBUȚII PRIVIND METODELE ȘI TEHNICILE DE ANALIZĂ BIG DATA PENTRU CREȘTEREA CALITĂȚII PROCESELOR EDUCAȚIONALE

Domeniul Calculatoare și Tehnologia Informației

Conducător științific,
conf. univ. dr. ing. MIRELA DANUBIANU

Doctorand,
DANIELA CODREANU (MARCU)

Suceava, 2023



Proiect cofinanțat din Fondul Social European prin Programul Operațional Capital Uman 2014-2020

Această lucrare a beneficiat de suport financiar prin proiectul

***Excelență academică și valori antreprenoriale - sistem de burse
pentru asigurarea oportunităților de formare și dezvoltare a
competențelor antreprenoriale ale doctoranzilor și
postdoctoranzilor – ANTREPRENORDOC***

Cod Proiect: SMIS 123847

Cod Contract: POCU/380/6/13

Axa prioritară 6 - Educație și competențe

Prioritatea de investiții – Îmbunătățirea utilității sistemelor de educație și formare pentru piața muncii, facilitarea trecerii de la educație la muncă și consolidarea sistemelor de educație și formare profesională și a calității lor, inclusiv prin mecanisme pentru anticiparea competențelor, adaptarea programelor de învățământ și crearea și dezvoltarea de sisteme de învățare bazate pe muncă, inclusiv sisteme de învățare duale și de ucenicie.

Apel de proiecte POCU/380/6/13/ „Sprijin pentru doctoranzi și cercetători post-doctorat”

Beneficiar: Parteneri:



Universitatea
Ștefan cel Mare
Suceava



ICECON S.A.

INSTITUTUL DE CERCETĂRI PENTRU ECHIPAMENTE ȘI TEHNOLOGII ÎN CONSTRUCȚII
RESEARCH INSTITUTE FOR CONSTRUCTION EQUIPMENT AND TECHNOLOGY



CAMERĂ DE COMERȚ, INDUSTRIE,
NAVIGATIE ȘI AGRICULTURĂ CONSTANȚA



Cuprins

Cuvinte cheie	5
1. Introducere	6
2. Stadiul actual al cercetărilor – o analiză a literaturii de specialitate	8
2.1. Scopul și obiectivele analizei literaturii de specialitate. Întrebarea de cercetare	8
2.2. Procesul de căutare și selecție	8
2.2.1. Stabilirea criteriilor de includere și excludere	8
2.2.2. Descrierea strategiei de căutare pentru obținerea rezultatelor primare	9
2.2.3. Descrierea procedurilor de selecție intermediară și finală	10
2.3. Sinteza rezultatelor. Concluzii	10
3. Big Data	12
3.1. Conceptul Big Data	12
3.2. Soluții arhitecturale pentru stocarea și procesarea Big Data	12
3.2.1. Ecosistemul Hadoop	12
3.2.2. Soluții cloud pentru Big Data	12
3.2.3. Analiză comparativă a serviciilor AWS și Google Cloud.....	13
3.3. Metode, tehnici și instrumente de analiză Big Data	14
3.4. Tehnici specifice prelucrării limbajului natural și imaginilor	14
4. Colectarea datelor utilizate în cercetare. Seturi de date	14
4.1. Setul de date <i>EduBac</i>	14
4.2. Setul de date <i>claritin_october_twitter_side_effects</i>	14
4.3. Setul de date <i>191-opinii-despre-școală</i>	15
4.4. Setul de date <i>Opinii</i>	15
4.5. Setul de date <i>Imagini-generale</i>	16
4.6. Setul de date <i>Imagini-Edu</i>	16
5. Cercetări experimentale privind utilizarea soluțiilor de analiză Big Data pentru creșterea calității proceselor educaționale	17
5.1. Utilizarea unei arhitecturi Hadoop Single Node Cluster 3.3.0 pe o mașină Ubuntu 20.04 LTS pentru prelucrarea datelor de tip text	17
5.2. Studiu de caz privind utilizarea Learning Analytics pentru analiza comportamentului și tipologiei utilizatorilor de site-uri educaționale.....	19
5.3. Investigație privind selecția tehnicilor de clasificare pentru predicția rezultatelor la Bacalaureat .	21
5.3.1. Metodologie de lucru	21
5.3.2. Rezultate obținute. Concluzii.....	22

5.4.	Implementarea proiectelor de analiză a sentimentelor folosind serviciul AutoML Language din Google Cloud	28
5.4.1.	Modelul ASG	28
5.4.2.	Modelul Opinii	30
5.4.3.	Comparație între modelele ASG și Opinii. Concluzii	31
5.5.	Analiza sentimentelor cu modelele ASO - Ekman și ASO – Plutchik	32
5.5.1.	Emoțiile. Modelele Ekman și Plutchik.....	32
5.5.2.	Modelele ASO-Ekman și ASO-Plutchik.....	32
5.6.	Proiectarea și implementarea proceselor de analiză a sentimentelor din imagini	33
5.6.1.	Analiza sentimentelor folosind AutoML Vision din Google Cloud	33
5.6.2.	Analiza sentimentelor folosind platforma Clarifai.....	35
5.6.3.	Recomandări privind alegerea soluțiilor adecvate analizei sentimentelor elevilor referitoare la procesul educațional.....	37
6.	Recomandări privind direcțiile de acțiune în domeniul educației.....	38
7.	Concluzii, contribuții personale și direcții viitoare de cercetare.....	39
7.1.	Concluziile tezei.....	39
7.2.	Contribuții personale	40
7.2.1.	Contribuții teoretice	40
7.2.2.	Contribuții practice.....	41
7.3.	Diseminarea rezultatelor	42
7.3.1.	Conferințe internaționale	42
7.3.2.	Lista articolelor publicate în jurnale și proceeding-uri ale conferințelor	43
7.3.3.	Implicarea în proiecte de cercetare	44
7.4.	Direcții viitoare de cercetare	44
	Acknowledgment.....	45
	Bibliografie.....	46

Cuvinte cheie

BIG DATA, EDUCATIONAL DATA MINING, LEARNING ANALYTICS, CLOUD, DATA MINING, MACHINE LEARNING, EDUCATION, DATA, NATURAL LANGUAGE PROCESSING, COMPUTER VISION, SENTIMENT ANALYSIS, ALGORITHMS, EVALUATION METRICS, LITERATURE REVIEW, HADOOP, DEEP NEURAL NETWORK, CONVOLUTIONAL NETWORK, CLARIFAI, KNN, ADABOOST, SVM, NAIVE BAYES, ORANGE, AUTOML, EKMAN, PLUTCHIK

1. Introducere

Nevoia de schimbare a paradigmelor educaționale se vede în conținuturi, servicii, tehnologii, mentalități. Planurile cadru și programele școlare din învățământul preuniversitar au suferit modificări de formă și mai puțin de fond, în ultimii 10-15 ani. Există discipline, așa cum este informatica, cu o dinamică specială, care au nevoie frecvent de curriculum nou, adaptat prezentului. O soluție viabilă este crearea și utilizarea cursurilor opționale, împreună cu mărirea numărului de ore alocate acestora în proiectarea curriculum-ului la decizia școlii.

Pe de altă parte, implementarea Strategiei de Digitalizare a Educației din România 2021-2027 SMART-Edu, care are ca scopuri crearea unui ecosistem digital adaptabil dinamicii schimbărilor și crearea de resurse educaționale deschise, va genera o cantitate foarte mare de date. Aceasta crează oportunități și direcții noi de cercetare pentru domenii precum Big Data, Data Mining, Machine Learning, cloud computing.

În prezent, tendința companiilor și comunităților educaționale internaționale de a lua decizii care se bazează pe date, duce la o creștere a necesității dezvoltării unor noi soluții de procesare a Big Data. Aceasta ridică valoarea datelor dar și probleme importante de etică și legislație.

Explorarea și analiza datelor ce provin din activitățile de învățare au condus la apariția unor domenii noi de cercetare precum Learning Analytics și Educational Data Mining. Acestea pot contribui la creșterea calității proceselor educaționale prin studierea și înțelegerea comportamentului și rezultatelor elevilor analizând datele care provin din interacțiunile acestora cu mediile de învățare.

Chiar dacă efortul de a crea platforme și instrumente noi, care înglobează servicii moderne de analiză, sisteme de recomandări, clasificarea utilizatorilor pentru livrare de conținut și evaluare particularizată și altele nu produce profit imediat, poate avea consecințe deosebit de favorabile privind educația pe termen lung. Studierea și cunoașterea, în timp real, a feed-back-ului elevilor privind activitățile educative și dinamica psiho-comportamentală (gândire, emoție, percepție, motivație) a participanților la actul de învățare pot contribui categoric la creșterea calității proceselor de instruire.

Scopul cercetărilor efectuate este de a identifica modalități de utilizare a tehnologiilor de prelucrare a Big Data pentru creșterea calității serviciilor din educație, în special din sistemul preuniversitar.

Obiectivele propuse sunt:

(O₁) Realizarea unei analize sistematice și critice asupra stadiului actual al cercetărilor

(O₂) Colectarea datelor din diverse surse ale sistemului educațional și constituirea seturilor potrivite pentru analiză

(O₃) Realizarea de cercetări experimentale privind alegerea și utilizarea soluțiilor de analiză Big Data pentru creșterea calității serviciilor din domeniul educațional

(O₄) Elaborarea unor recomandări privind direcții viitoare de acțiune pe baza rezultatelor analizelor efectuate.

Lucrarea de doctorat conține 7 capitole și bibliografie: Introducere, Stadiul actual al cercetărilor – o analiză a literaturii de specialitate, Big Data, Colectarea datelor utilizate în cercetare. Seturi de date, Cercetări experimentale privind utilizarea soluțiilor de analiză Big Data pentru creșterea calității proceselor educaționale, Recomandări privind direcțiile de acțiune în domeniul educației, Concluzii, contribuții personale și direcții viitoare de cercetare.

Teza conține și două anexe: Google Cloud. Scurtă descriere a principalelor operații necesare configurării serviciului AutoML Language cu scopul utilizării modelelor ASG și Opinii pentru

analiza sentimentelor și Automatizarea procesului de splitare a fișierului Excel în fișiere TXT folosind limbajul de programare C++.

Stadiul actual al cercetărilor – o analiză a literaturii de specialitate

Pentru a obține o imagine a stadiului actual al cercetărilor și preocupărilor din domeniul Big Data din educația românească, am efectuat o analiză a literaturii de specialitate (Capitolul 2). Căutarea și selecția studiilor a fost un proces sistemic de analiză în 4 baze de date: IEEE Xplore, ScienceDirect, Web of Science și Google Scholar. În urma aplicării criteriilor de includere și excludere, au fost selectate 38 de articole publicate în jurnale, proceeding-uri ale conferințelor sau capitole de carte. Formularea concluziilor s-a făcut printr-o abordare calitativă, meta-analitică și meta-narativă.

Big Data

Capitolul 3 al lucrării cuprinde aspecte teoretice privind: conceptul Big Data, soluții arhitecturale pentru stocarea și procesarea Big Data, analizele Big Data, metode de clasificare și grupare a datelor, Learning Analytics și Educational Data Mining, precum și prelucrarea limbajului natural, analiza sentimentelor și Computer Vision, rețele neuronale cu structură profundă și rețele neuronale convoluționale.

Colectarea datelor utilizate în cercetare. Seturi de date

Seturile de date pe care le-am utilizat în cercetările efectuate sunt originale și provin din diferite surse din educație. În capitolul 4 am prezentat maniera de colectare și sursele de unde provin datele, precum și structurarea și alegerea caracteristicilor în funcție de scopul urmărit. Pentru a efectua analiza sentimentelor din text, am conceput un sistem de etichetare, în care au fost implicați doi profesori, doi elevi și un student. Metoda și rezultatele sunt prezentate în acest capitol. O parte din seturile de date vor fi publicate în regim open acces, după susținerea publică a lucrării de doctorat.

Cercetări experimentale privind utilizarea soluțiilor de analiză Big Data pentru creșterea calității proceselor educaționale

Cercetările experimentale efectuate sunt contribuții originale privind: analiza sentimentelor din date provenite din educație, clasificarea emoțiilor de bază folosind modelele psihologilor Ekman și Plutchik, predicții asupra rezultatelor obținute de elevi la examenul de Bacalaureat, crearea unei arhitecturi Hadoop Single Node Cluster pe o mașină Ubuntu cu scopul de a utiliza componenta MapReduce la procesarea textului, utilizarea unui instrument de analitica datelor pentru a evidenția importanța acestuia în identificarea tipurilor de utilizatori și analiza interacțiunilor acestora cu platformele de învățare pentru a eficientiza experiența de utilizare.

Recomandări privind direcțiile de acțiune în domeniul educației

Capitolul conține 6 direcții de acțiune pe care le-am identificat ca urmare a cercetărilor teoretice și experimentale efectuate.

Concluzii, contribuții personale și direcții viitoare de cercetare

Capitolul prezintă: concluziile tezei, procesele și metodele de atingere a obiectivelor stabilite, perspective noi de cercetare, contribuțiile teoretice și practice și diseminarea rezultatelor.

Lucrarea se încheie cu două anexe și referințe bibliografice.

2. Stadiul actual al cercetărilor – o analiză a literaturii de specialitate

Cu scopul de a cerceta progresele și interesul pentru domeniul aplicării metodelor, tehnologiilor și arhitecturilor specifice pentru Big Data în educație, a fost considerată oportună realizarea unei analize sistematice și critice a literaturii de specialitate care vizează autorii români și date, instrumente, cadre de lucru, aplicații provenite sau construite pentru componente ale sistemului de educație din România.

2.1. Scopul și obiectivele analizei literaturii de specialitate. Întrebarea de cercetare

Scopul efectuării analizei literaturii de specialitate este *obținerea de informații privind utilizarea arhitecturilor, instrumentelor, analizelor și tehnologiilor specifice Big Data în sistemul educațional românesc, în special în cel preuniversitar*. Totodată analiza trebuie să permită nu doar colectarea sistematică a informațiilor, cât și sintetizarea rezultatelor. Acest lucru va fi util pentru a identifica direcții noi de cercetare sau pentru a aduce contribuții la completarea unor studii anterioare. Folosind rezultatele analizei literaturii de specialitate se întrevide posibilitatea de a utiliza resurse și instrumente existente cu scopul de a verifica, în manieră comparativă, oportunitatea utilizării lor pe scară largă, cu costuri cât mai reduse, de către cercetători sau cadre didactice, nu neapărat cu experiență în domeniul analiticii datelor sau altele conexe cu acesta. În acest fel există o șansă în plus ca, prin utilizarea noilor tehnologii informaționale, să se aducă contribuții la creșterea calității serviciilor din domeniul educației.

Având în vedere acestea, am stabilit următoarele **obiective**:

O₁: *Evaluarea stadiului actual al cercetărilor privind aplicarea tehnicilor specifice explorării datelor din domeniul educațional românesc;*

O₂: *Identificarea principalelor proceduri, cadre de lucru, arhitecturi, instrumente, moduri de analiză a datelor, în special a Big Data, folosite de autori români sau străini pentru experimente desfășurate cu elevi, profesori și alte categorii de personal din sistemul educațional românesc;*

O₃: *Prezentarea unor direcții viitoare de cercetare.*

2.2. Procesul de căutare și selecție

Procesul de căutare și selecție ale informațiilor necesare analizei literaturii de specialitate respectă metodologia căutării sistematice a lui Kitchenham [7] ce presupune parcurgerea etapelor: planificarea, efectuarea și raportarea căutării.

2.2.1. Stabilirea criteriilor de includere și excludere

Selecția rezultatelor primare are la bază criteriile de includere și de excludere. Pentru stabilirea acestora, am definit un **criteriu general de selecție**, în acord cu întrebarea de cercetare: *studii care implică realizarea de: cadre de lucru, aplicații, experimente, analize a literaturii de specialitate, modele de bună practică, dezvoltarea unor algoritmi de explorare a datelor sau utilizarea unor instrumente și arhitecturi specifice Big Data pentru sistemul educațional românesc.*

Studiile selectate trebuie să respecte unul sau mai multe din **criteriile de includere**:

I₁: *studii care prezintă o analiză a literaturii de specialitate ce vizează aspecte ale Big Data sau Educational Data Mining din sistemul educațional românesc;*

I₂: studii care prezintă cadre de lucru, metode de explorare a datelor, analiza datelor din sistemul educațional românesc;

I₃: studii care descriu utilizarea de instrumente, tehnologii, aplicații și arhitecturi specifice Big Data pentru rezolvarea de probleme specifice învățământului românesc;

I₄: studii care au autori români și/sau prezintă rezolvarea unor probleme sau studii de caz, algoritmi de explorare a datelor, sisteme de recomandare, cadre de lucru, aplicații construite pentru sistemul educațional, nu neapărat românesc.

Criteriile de excludere:

E₁: studiile care nu sunt din perioada 2008-2018;

E₂: studiile care nu sunt relevante pentru întrebarea de cercetare;

E₃: studiile care nu au autori români.

În selecție vor fi incluse: articole de cercetare, capitole de carte, editoriale, recenzii de carte, analize a literaturii de specialitate, rapoarte tehnice și proceeding-uri ale lucrărilor prezentate la conferințe.

2.2.2. Descrierea strategiei de căutare pentru obținerea rezultatelor primare

Stabilirea expresiilor de căutare pentru fiecare bază de date s-a făcut prin rafinarea iterativă a căutării pentru diferiți termeni și expresii. Pentru o căutare avansată am utilizat operatori de tip boolean. În tabelul 1 sunt prezentate expresiile de căutare pentru fiecare bază de date interogată și numărul de rezultate returnate în faza inițială.

Tabel 1. Distribuția numărului de studii selectate în etapa inițială pentru fiecare bază de date și expresiile de căutare cu care au fost obținute rezultatele

Baza de date	Expresia de căutare	Nr. studii selectate în urma aplicării expresiei de căutare
Google Scholar	Romanian Big Data educational data mining methods learning OR students OR class OR teachers OR methods - medicine -marketing -health -media	173
Web of Science	Expresia de căutare în forma prescurtată: ((ALL=(Romanian)) AND (ALL=(Big Data methods)) OR ALL=((Educational Data Mining)))	64
IEEE Xplore	((("Full Text Only":Big Data methods OR ("Full Text Only":EDM OR "Full Text Only":Educational Data Mining)) AND ("Full Text Only":learning OR "Full Text Only":students OR "Full Text Only":class OR "Full Text Only":teachers OR "Full Text Only":methods)) AND "Full Text Only":Romanian)	388
Science Direct	Romanian AND "Big Data methods" OR "educational data mining" -medicine -marketing -health -media	131

2.2.3. Descrierea procedurilor de selecție intermediară și finală

Înainte de selecția studiilor pentru efectuarea analizei propriu-zise, a existat o etapă intermediară. În această fază, am parcurs titlurile, autorii și, eventual, numele jurnalului, a cărții sau a conferinței și am selectat acele rezultate care aveau o șansă să îndeplinească unul din criteriile de includere. Pentru selecția finală, a fost necesară parcurgerea titlurilor, autorilor, rezumatelor, cuvintelor cheie, editorilor, titlurilor jurnalelor și conferințelor, a domeniilor, în cazul în care acestea sunt specificate și, dacă toate acestea nu au fost concludente, am citit articolele în întregime. Rezultatele acestor etape sunt prezentate în tabelul 2.

Tabel 2. Distribuția numărului de studii selectate în fiecare etapă pentru fiecare bază de date

Baza de date	Număr de studii selectate în urma aplicării expresiei de căutare (selecția primară)	Număr de studii rezultate în urma selecției intermediare	Număr de studii selectate în urma aplicării criteriilor de includere și excludere (selecția finală)
Google Scholar	173	18	4
Science Direct	131	94	2
Web of Science	64	38	28
IEEE Xplore	388	40	4

Selecția studiilor în etapa finală a fost făcută prin aplicarea criteriilor de includere și excludere.

2.3. Sinteza rezultatelor. Concluzii

Analiza literaturii de specialitate din perioada 2008-2018 a fost realizată folosind o căutare sistematică în 4 baze de date: Google Scholar, IEEE Xplore, Web of Science și ScienceDirect. Acestea au fost alese deoarece sunt printre cele mai relevante pentru domeniul Computer Science. În urma aplicării expresiilor de căutare, au fost selectate 756 de rezultate. O parcurgere a titlurilor, detaliilor despre autori, rezumatelor a condus la o nouă selecție astfel încât, în etapa intermediară au rămas 190 de lucrări. În final, am selectat 38 de studii din 26 de domenii/discipline: Education, E-Learning, Big Data, Computer Science, Artificial Intelligence, Computer aided instruction, Scientific information systems, Interdisciplinarity Applications, Engineering, Educational Data Mining și altele pentru care am efectuat o evaluare folosind o abordare calitativă, meta-analitică și meta-narativă.

Parcurgerea articolelor în vederea selecției finale a scos la iveală faptul că interesul cercetătorilor români privind problematica Big Data din educație, analitica învățării, explorarea datelor, folosirea noilor tehnologii, care s-au dovedit extrem de utile pentru îmbunătățirea serviciilor din alte domenii, este cu mult mai scăzut decât al cercetătorilor din alte țări. Spre exemplu, am putut identifica doar 4 lucrări care prezintă cercetări privind explorări ale Big Data

din educație care satisfac criteriul de includere I₃. Absența arhitecturilor specifice pentru stocarea și procesarea Big Data, costurile foarte mari ale serviciilor din cloud pentru date de mari dimensiuni, accesul limitat la instrumente și cadre de lucru specifice explorării datelor, lipsa de specialiști calificați în domeniile vizate de întrebarea de cercetare, mai ales în sistemul preuniversitar ar putea fi câteva motive pentru care în literatura de specialitate din perioada studiată nu am identificat foarte multe articole ale autorilor români, care privesc educația din România. Procesul de digitalizare a serviciilor este unul foarte lent, ceea ce face din colectarea datelor un proces foarte anevoios. La aceasta se adaugă și politicile privind datele cu caracter personal. În plus, profesorii nu sunt motivați să se specializeze în domenii noi, să inoveze, să identifice surse de finanțare, să efectueze cercetări a căror finalitate poate să nu fie cea așteptată.

În sprijinul acestor concluzii vin și observațiile: majoritatea studiilor privesc învățământul universitar și aproximativ un sfert din rezultate sunt cercetări la nivel teoretic.

Niciun studiu nu prezintă o analiză a literaturii de specialitate care să vizeze aspecte ale Big Data sau Educational Data Mining din sistemul educațional românesc, conform criteriului de includere I₁. O direcție viitoare de cercetare ar putea fi realizarea unei analize a literaturii de specialitate din altă perioadă de timp, după 2015, printr-o căutare și în alte baze de date, dar și în jurnale, cărți, recenzii, rapoarte de cercetare și altele.

Studiile selectate pentru analiză au ca scop eficientizarea proceselor academice prin:

- identificarea cursanților cu risc de abandon școlar;
- crearea profilurilor studenților, gruparea și clasificarea acestora;
- sprijinirea deciziilor manageriale prin explorarea datelor;
- corelarea traiectoriilor educaționale ale elevilor cu activitățile de pe piața muncii;
- predicții privind performanțele studenților;
- preprocesarea și gruparea datelor după funcționalitate: administrație, cursuri, evaluare, etc;
- ajustarea curriculum-ului;
- construirea de sisteme de recomandări de conținut și livrarea de bibliografie.

Datele utilizate au fost colectate prin chestionare sau din activități online, fișiere de tip jurnal, forumuri de discuții, click-stream, interacțiunile cu platforme de învățare precum MOOC, PULSE, LMS, Tesys.

Instrumentele și cadrele de lucru utilizate în realizarea experimentală a cercetărilor sunt: WEKA, Rapid Miner, Le ProVe-uCluster, WAT, TAALES, TAACO, Reader Bench, SEANCE, Web Django, Flask.

Pentru implementarea algoritmilor folosiți în clasificarea și preprocesarea datelor, prelucrarea Big Data și vizualizare au fost folosite o serie de limbaje precum: Python (plus pachetul Scikit-Learn), Scala, JavaScript, HTML 5.

Pentru gestionarea lucrului cu Big Data a fost folosit sistemul de procesare Spark.

Pentru clasificarea datelor sunt utilizați: arborii de decizie, regulile de asociere, perceptronul multistrat și rețele neuronale. Cea mai utilizată tehnică de grupare a datelor este K-Means. O singură cercetare folosește algoritmi bazați pe ontologie.

Algoritmii utilizați de autorii cercetărilor selectate pentru analiză sunt: Singular Value Decomposition, algoritmi de eșantionare, Linear Discrimination Analysis, FarthestFirst, C4.5, Large Margin Nearest Neighbor Regression, Decision Tree CART, Extra Tree Classifier, Logistic Regression, C-Support Vector Classification, Random Forest Classifier.

Având în vedere analiza efectuată, am putut identifica câteva direcții noi de cercetare:

- utilizarea și crearea de cadre și instrumente de lucru pentru explorarea datelor;
- utilizarea serviciilor cloud pentru efectuarea operațiilor de colectare, stocare și analiză specifice Big Data pentru sistemul educațional;

- utilizarea tehnicilor potrivite pentru analiza și modelarea datelor;
- crearea de platforme educaționale moderne pentru sistemul preuniversitar care folosesc instrumente ce au fost experimentate cu succes în alte domenii de activitate.

Contribuția acestei analize sistematice și critice a literaturii de specialitate este formarea unei priviri de ansamblu, cât mai apropiate de realitate, în legătură cu cercetările efectuate în domeniul prelucrării Big Data din sistemul educațional românesc. Nu au fost identificate studii asemănătoare în procesul de căutare sistematică, de aceea sinteza rezultatelor dezvăluie noi direcții de cercetare în acest domeniu și chiar oportunitatea realizării unei analize a literaturii prin explorarea mai multor baze de date, pentru o altă perioadă de timp.

3. Big Data

3.1. Conceptul Big Data

Termenul de Big Data se referă la colecții imense de date, de o mare varietate, care necesită instrumente, tehnologii și infrastructuri speciale pentru a fi stocate și prelucrate cu scopul de a extrage informații utile din acestea, cunoștințe ce pot fi furnizate către factorii de decizie prin intermediul analizelor Big Data [210].

Paradigma 5V se referă la 5 dintre cele mai cunoscute caracteristici ale Big Data. Acestea sunt: volumul, varietatea, viteza, veridicitatea și valoarea. În prezent sunt cunoscute 56 de caracteristici printre care pot fi enumerate: variability, vocabulary, volatility, visualization, viscosity, vendible, visible, violable și altele [216].

3.2. Soluții arhitecturale pentru stocarea și procesarea Big Data

3.2.1. Ecosistemul Hadoop

Arhitectura generală a unui ecosistem Hadoop cuprinde mai multe componente: Hadoop Distributed File System (HDFS, sistem distribuit pentru organizarea fișierelor), Yet Another Resource Negotiator (YARN, cadru utilizat pentru planificarea și împărțirea gestionării resurselor în cluster), MapReduce (sistem pentru procesarea paralelă a sarcinilor), Common (cadru pentru utilitățile comune acceptate de celelalte module), Ozone (magazin de obiecte) [219].

Pe lângă acestea, sistemului Hadoop îi pot fi asociate și alte componente precum: Mahout (bibliotecă creată pentru algoritmi de învățare automată și data mining, compatibili cu MapReduce), Hive (platformă de tip data warehouse utilizată pentru gestionarea Big Data stocate în HDFS cu ajutorul unui limbaj similar cu SQL), Pig (limbaj cu compilator care produce sarcini MapReduce pentru procesarea Big Data) și altele [220].

Există două tipuri de arhitecturi Hadoop, ambele având la bază același principiu de organizare într-o structură ierarhică care conține unul sau mai multe NameNodes și DataNodes: Single Node Cluster și Multiple Node Cluster.

3.2.2. Soluții cloud pentru Big Data

Există 3 tipuri de cloud: public (infrastructura de calcul se află la furnizor), privat (aparține unei organizații, având infrastructura la sediul organizației sau la furnizor) și hibrid (aplicațiile esențiale sunt găzduite pe serverele clientului, iar restul la locația furnizorului). Un motiv pentru

alegerea unui cloud hibrid este legislația în materie de protecție a datelor personale; unele date nu pot părăsi țara, spre a fi stocate în alte regiuni disponibile prin serviciile cloud. În acest domeniu, Amazon vine cu soluția AWS Outposts disponibilă cu un rack 42U ce poate scala până la 96 rack-uri care furnizează infrastructura fizică și logistică pentru a utiliza toate serviciile AWS pentru procesare locală a Big Data [226].

Principalele servicii din cloud sunt: IaaS (Infrastructure as a Service), PaaS (Platform as a Service), SaaS (Software as a Service) [227].

3.2.3. Analiză comparativă a serviciilor AWS și Google Cloud

Atunci când este necesară luarea unei decizii privind cea mai potrivită platformă de cloud trebuie avute în vedere avantajele și dezavantajele fiecăreia raportate la scopul propus. Este necesară așadar o analiză comparativă. Comparatia este elaborată având în vedere serviciile, dar și aspecte precum: zona de disponibilitate, securitate, modele de prețuri, rețea și altele.

Google Cloud oferă peste 60 de servicii în cloud, pe când Amazon peste 200, lucru explicabil având în vedere experiența de piață mai mare cu 5 ani a AWS [233].

Pentru stocare, Amazon oferă S3, iar Google serviciul Cloud Storage. Ambele companii oferă soluții scalabile, durabile și cu o disponibilitate ridicată, care acoperă stocarea obiectelor cu dimensiuni de până la 5TB, block-urilor și a sistemelor de fișiere. Alegerea unuia dintre acestea se face cunoscând beneficiile oferite și modalitățile de tarifare.

Pentru procesare, Amazon oferă serviciul EC2, iar Google Compute Engine. Ambele alternative sunt ușor de utilizat, oferă control complet asupra resurselor de procesare și securitate, precum și o serie de instrumente utile în realizarea aplicațiilor la nivel de IaaS, PaaS, containere și Serverless Functions (serviciul Lambda de la AWS, respectiv Google Cloud Functions) [233].

Privitor la serviciile pentru baze de date, ambele companii oferă posibilitatea prelucrării și stocării datelor RDBMS (Relational Database Management System) – care sunt la baza sistemelor moderne de baze de date precum: Microsoft SQL Server, Oracle, MySQL și altele – și NoSQL. Din nou, utilizarea unui serviciu ține cont de costuri, de tipul datelor, etc.

Atât AWS cât și Google oferă puncte de prezență în locații din întreaga lume.

Ambele companii oferă servicii pentru analitica Big Data și pentru învățare automată.

Pentru securitate, AWS și Google Cloud oferă servicii implicite foarte sigure care protejează datele stocate în cloud. Totodată, ambele companii oferă și servicii suplimentare de securitate. Pentru gestionarea identităților digitale, AWS și Google pun la dispoziție serviciul Identity and Access Management.

Pentru vizualizarea datelor, AWS oferă serviciul AWS QuickSight, cu taxare pe sesiune, pe când Google pune la dispoziție, gratuit, serviciul Google Data Studio ce poate fi utilizat la generarea de rapoarte folosind datele stocate în cloud.

Pentru realizarea cercetărilor experimentale, am optat pentru Google Cloud deoarece: oferă utilizarea gratuită, pe o perioadă de 30 de zile, în prezent (anterior fiind de 90 de zile) a majorității serviciilor, inclusiv a celor pentru învățarea automată, precum AutoML Language sau AutoML Vision.

Contribuția acestui studiu comparativ constă în facilitarea alegerii celei mai convenabile variante de cloud pentru efectuarea cercetărilor experimentale privind clasificarea textelor și a imaginilor, printr-o analiză teoretică dar și practică a serviciilor oferite de cele două companii.

3.3. Metode, tehnici și instrumente de analiză Big Data

Sunt prezentate aspecte teoretice privind analizele Big Data, EDM și LA.

În această secțiune a lucrării sunt descrise, de asemenea, metode de clasificare și regresie: K-Nearest Neighbors, AdaBoost, Support Vector Machine, Naïve Bayes și metoda de grupare K-Means. Acestea au fost selectate spre prezentare deoarece sunt folosite în cercetările experimentale efectuate.

3.4. Tehnici specifice prelucrării limbajului natural și imaginilor

Secțiunea conține prezentarea, la nivel teoretic, a unor noțiuni precum: prelucrarea limbajului natural, analiza sentimentelor, computer vision, rețele neuronale cu structură profundă și rețele neuronale convoluționale.

4. Colectarea datelor utilizate în cercetare. Seturi de date

4.1. Setul de date *EduBac*

Datele au fost colectate folosind aplicația informatică dedicată examenului de Bacalaureat, printr-o operație de exportare a rezultatelor în fișiere Excel. Pentru fiecare an din perioada 2012-2019, am obținut câte un fișier. Datele au fost unificate într-un singur fișier, eliminând informații privitoare la datele personale ale elevilor, sălile de examen, și altele. Rezultatele provin de la Colegiul Economic "Dimitrie Cantemir" din Suceava.

Pentru utilizarea datelor, acestea au trecut printr-o operație de preprocesare în care am verificat dacă înregistrările sunt complete. Am validat 2245 de instanțe și 13 caracteristici: Proba 1, Proba 2, Proba 3, Media, Statut (variabila țintă), An, Sex, Specializarea, Profil, Competențe digitale, Competențe lingvistice, Proba D, Promoția curentă.

4.2. Setul de date *claritin_october_twitter_side_effects*

Acest set de date este alcătuit din tweet-uri, din luna octombrie a anului 2012, în care este menționat cuvântul *claritin*. Tweet-urile sunt etichetate cu *sentiment* care are valori în intervalul [1,5]. Acesta a fost adăugat pe platforma *data.world* de către Dave Oleson la 13 noiembrie 2013 și conține 4900 de înregistrări [299].

Datele utilizate în cercetarea efectuată au fost descărcate de pe platforma *data.world*, în urma efectuării unor interogări SQL, cu scopul de a memora în format CSV tweet-urile pe categorii de sentimente, având ca rezultat fișiere de forma: **s- claritin-twitter-QueryResult.CSV**, unde *s* este o valoare între 1 și 5. Exemplu:

```
SELECT content FROM  
`1384367161_claritin_october_twitter_side_effects_1`where sentiment=1
```

Pentru realizarea cercetărilor, datele au fost organizate sub forma unei arhive ZIP, *documents.zip*. Pentru a crea această structură de foldere și fișiere a fost necesară splitarea fiecărui fișier CSV descărcat de pe platforma *data.world* în fișiere de tip text astfel încât fiecare înregistrare din fișierul CSV să fie salvată într-un fișier de tip .TXT separat. Am automatizat această operație prin scrierea unui program în limbajul de programare C++. Pentru compilarea și executarea

proiectului am utilizat mediul de programare open source Code::Blocks 20.03 și compilatorul GNU GCC Compiler.

A rezultat un set de date, alcătuit din 4635 de fișiere.

4.3. Setul de date 191-opinii-despre-școală

Pentru colectarea datelor, am aplicat un chestionar unui număr de 196 de respondenți, în perioada 7 – 20 februarie 2020. Formularul Google cu titlul *Chestionar "Opinii despre școală"* trimis spre completare elevilor de clasa a 11-a și a 12-a de la profilul matematică-informatică, din zona Sucevei. Am selectat răspunsurile la întrebarea: *Te rog să descrii în câteva cuvinte, cu sinceritate, care sunt opiniile tale despre școală* și le-am salvat într-un fișier CSV, în câmpul *Opinii*. Textul a fost apoi tradus din română în engleză și a rezultat un nou fișier CSV. Setul de date *Opinii despre școală* este constituit din: fișierul *opinii-despre-scoala-engleza.CSV*, fișierul *opinii-despre-scoala-engleza.PDF* (aceleași date, dar salvate cu altă extensie de fișier) și folderul *opinii-engleza* care conține 191 de fișiere de tip TXT, fiecare având drept conținut câte o opinie despre școală, dintre cele care au fost validate.

4.4. Setul de date Opinii

Pentru alcătuirea acestui set de date am parcurs mai multe etape: alegerea indicatorilor ce vizează formularea opiniei despre școală, realizarea și distribuirea chestionarului *Opinii despre școală*, etichetarea manuală a opiniilor și constituirea setului de date.

Pentru a obține răspunsuri unitare, care să vizeze aceleași aspecte generale privind educația, respondenții au fost invitați să se ghideze după anumiți indicatori. Aceștia sunt: egalitatea de gen și de șanse, participarea la activități non-formale sau proiecte extrașcolare în anii de liceu, măsura în care educația privind drepturile omului, educația financiară, cea din punct de vedere al sănătății sunt integrate în curriculumul obligatoriu sau opțional, accesul la internet, la apă potabilă și căldură, la infrastructura adaptată persoanelor cu dizabilități, gradul de siguranță din școală referitor la relațiile cu colegii și profesorii, feed-back-ul profesorilor, gradul de satisfacție privind achizițiile dobândite în raport cu preocupările și interesele personale.

Am realizat chestionarul "Opinii despre școală" pe care l-am distribuit, în perioada 26 octombrie – 5 noiembrie 2021 unui număr de 1443 respondenți. Aceștia au fost selectați, pe bază de voluntariat, dintre elevii de liceu din România, din clasele a 9-a – a 13-a. Participarea a fost anonimă, fără accesarea de date personale de identificare precum: e-mail, telefon, adresă, nume, școală. Am obținut astfel un set valoros de date, pe care l-am utilizat în această lucrare, dar și în alte analize pe care le-am făcut.

Din totalul de 1443 de răspunsuri au fost validate 1272, acestea fiind utilizate în continuare, în urma unui proces de etichetare.

Pentru realizarea setului de date, am selectat răspunsurile deschise la întrebarea: *"Te rog să formulezi, în câteva cuvinte, opinia ta sinceră despre școală. În formularea acesteia, poți să te ghidezi după răspunsurile date la întrebările 4 - 10 sau să te referi la alte aspecte legate de școală"*. În continuare, fiecare răspuns necesită atribuirea unei clase de sentiment de la 0 la 4.

Pentru realizarea etichetării manuale a datelor, am format o echipă, în care am invitat încă un profesor de liceu, 2 elevi și un student la psihologie, în Olanda. Au fost atribuite scoruri de sentiment cuprinse între 0 și 4 (0 pentru cea mai nefavorabilă opinie) pentru fiecare opinie exprimată. Mesajele care conțineau doar caractere precum "..." sau doar cuvinte fără înțeles, de exemplu "blk" nu au fost validate.

Fiecare membru al echipei a atribuit propriile scoruri de sentiment. Mesajele care au primit scoruri diferite au fost trimise către etichetare celui de-al doilea profesor din echipă. La final, au fost analizate mesajele care au fost etichetate cu scoruri diferite de sentiment. Acestea au fost din nou citite și evaluate. Nu au existat mesaje etichetate cu 5 scoruri distincte, de aceea, scorul majoritar a fost cel atribuit în final mesajului evaluat.

Un exemplu de opinie etichetată cu scoruri diferite este: "There are no cases of inequality of opportunity, and the feedback of the teachers does not appear much because I am not very good at learning.". Aceasta a primit 1 de la un elev, 2 de la al doilea, 3 de la un profesor și 2 de la celălalt și 2 de la studentul la psihologie. Scor final: 2.

Am obținut 1272 de înregistrări asociate scorurilor de sentiment după cum urmează: 153 cu scorul 0, 155 cu scorul 1, 248 cu scorul 2, 222 cu scorul 3 și 494 cu scorul 4.

4.5. Setul de date *Imagini-generale*

Setul de date cuprinde 153 de imagini, în format .JPG, pe care le-am etichetat cu scopul de a identifica emoțiile: *fericire, încredere, teamă, tristețe*. Acestea conțin: persoane, peisaje din diferite anotimpuri, animale, obiecte, scris de mână și tipărit. Imaginile sunt personale iar etichetarea lor a fost făcută manual prin asocierea cu una din cele 4 clase.

Pentru realizarea unui experiment de analiză a sentimentelor folosind două etichete de clasă, pozitiv și negativ, am etichetat manual și împărțit datele în proporțiile: 80% pentru datele de antrenament, 10% pentru validarea modelului și 10% pentru datele de test.

4.6. Setul de date *Imagini-Edu*

Setul de date este alcătuit din 94 de imagini personale, obținute în timpul desfășurării de activități școlare și extrașcolare la Colegiul Național "Ștefan cel Mare" din Suceava și în timpul unor activități online, desfășurate la centrul meu de pregătire.



Figura 1. Imagini etichetate POZITIV

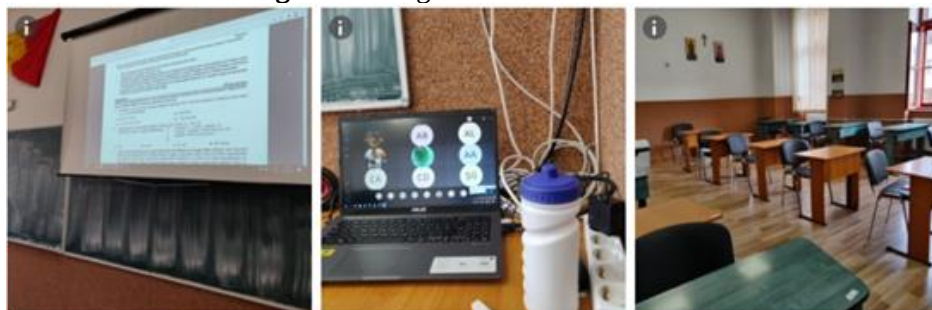


Figura 2. Imagini etichetate NEGATIV

Am etichetat manual aceste date, asociind 49 de imagini conceptului POZITIV și 45 conceptului NEGATIV (Fig.1, Fig. 2). Pentru efectuarea selecției am urmărit diversitatea situațiilor și contextelor în care se află elevii.

Contribuția adusă prin colectarea și constituirea seturilor de date pentru analiză este valorificată prin publicarea, în regim open-acces a acestora, spre a putea fi utilizată de oricine dorește să facă cercetări experimentale pe baza lor.

5. Cercetări experimentale privind utilizarea soluțiilor de analiză Big Data pentru creșterea calității proceselor educaționale

5.1. Utilizarea unei arhitecturi Hadoop Single Node Cluster 3.3.0 pe o mașină Ubuntu 20.04 LTS pentru prelucrarea datelor de tip text

Descrierea arhitecturii este prezentată în figura 3.

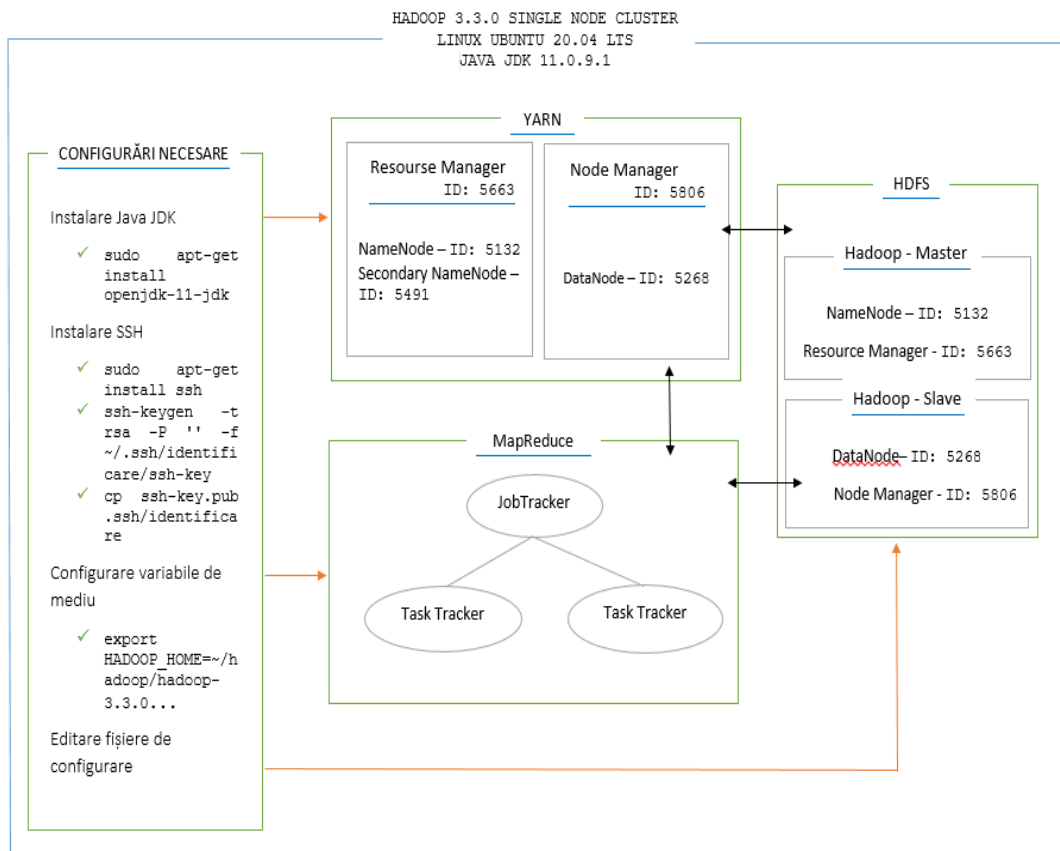


Figura 3. Schema principală a instalării unei arhitecturi Hadoop Single Node Cluster

Pentru efectuarea operațiilor cu fișiere, a fost utilizată componenta HDFS, pentru procesare, MapReduce și pentru gestionarea resurselor, YARN. Pentru verificarea integrității proceselor care se execută în nodurile clusterului Hadoop de pe mașina Ubuntu se utilizează instrumentul Java Virtual Machine Process Status (JVM).

Comanda JPS furnizează ID-ul și numele procesului pentru orice proces care rulează pe mașina virtuală Java, dacă nu sunt precizați parametri suplimentari. Așa cum se poate vedea în schema prezentată, fiecare proces are asociat un ID, de exemplu Resource Manager are atribuit ID-ul 5663. Odată ce s-a constatat că toate componentele YARN și HDFS sunt funcționale, pot fi efectuate sarcini de lucru.

Folosind un program, scris în Java, pentru MapReduce, din fișierul de intrare *propozitie.txt*, este preluat textul de analizat și copiat în HDFS fiind apoi procesat. Textul este împărțit în cuvinte. Pentru fiecare cuvânt, se determină frecvența de apariție. Ieșirile sunt salvate sub formă de perechi *cuvânt-număr de apariții* în folderul *wc_output* din HDFS. Experimentul poate fi dezvoltat, spre exemplu, în scopul efectuării analizei sentimentelor dintr-un text sau a creării unei aplicații de vizualizare a datelor, de tip word cloud, ce folosește frecvența de apariție a cuvintelor.

Odată lansată sarcina de lucru, MapReduce se conectează folosind portul 8032, la Resource Manager care alocă resursele necesare execuției sarcinii, pentru cele două etape: map și reduce. Spre exemplu, pentru job-ul curent au fost efectuate 8 operații de citire și 2 de scriere folosind HDFS totalizând 204 bytes, respectiv 108 bytes. Timpul total pentru mapare a fost de 2524 de milisecunde, iar pentru rezolvarea sarcinii reduce au fost necesare 2608 milisecunde. Cadrul MapReduce raportează: 2 înregistrări pentru *map input* și 16 pentru *map output*, 13 pentru gruparea intrărilor *reduce* și 13 pentru *reduce output*. S-au obținut, deci, 13 cuvinte, ordonate alfabetic având asociat numărul de apariții (Fig. 4).

```
danutza@danutza-All-Series:~/hadoop/hadoop-3.3.0/sbin$ hadoop fs -ls /wc_output
Found 2 items
-rw-r--r-- 1 danutza supergroup          0 2020-09-11 17:42 /wc_output/_SUCCESS
-rw-r--r-- 1 danutza supergroup       108 2020-09-11 17:42 /wc_output/part-r-00000
danutza@danutza-All-Series:~/hadoop/hadoop-3.3.0/sbin$ hadoop fs -cat /wc_output/part-r-00000
Hadoop 2
MapReduce 2
```

INPUT:

"învata Hadoop si MapReduce. Primul proiect MapReduce pentru Hadoop este usor daca se învata citind articole."

OUTPUT:

Hadoop 2
MapReduce 2
articole 1
citind 1
daca 1
este 1
învata 2
pentru 1
primul 1
proiect 1
se 1
si 1
usor 1

Figura 4. Rezultatul execuției sarcinii MapReduce

Pentru a procesa un text de 16 cuvinte, MapReduce a rulat sarcinile de lucru, pe mașina utilizată, în 5,232 secunde. Rezultatele pentru un text de aproximativ 60.000 de cuvinte, cam cât are această lucrare, ar fi obținute în aproximativ 5 ore și jumătate.

Contribuțiile aduse cu acest studiu constau în identificarea dificultăților și oportunităților de utilizare a arhitecturilor Hadoop pentru procesarea datelor provenite din sistemul educațional, dar și a unor direcții de acțiune necesare.

Având în vedere capacitatea computațională disponibilă în prezent în școli, pentru a crea arhitecturi pentru Big Data ar fi nevoie de investiții masive în infrastructură. Hadoop este o soluție care funcționează pe sisteme obișnuite, folosind distribuții Linux sau sistemul de operare Windows. Cu toate acestea, crearea arhitecturilor mai complexe, ce conțin componente suplimentare, este o

sarcină dificilă. Este necesară instruirea de specialiști în domeniu. În primă instanță, pot fi instruiți inginerii de sistem din școli.

Maximizarea valorii datelor colectate și procesate prin intermediul acestor arhitecturi are nevoie de conceperea unei strategii pentru dezvoltarea infrastructurii, mai întâi în școli pilot, instruirea și motivarea personalului și construirea cadrelor de reglementare privind politicile de prelucrare a datelor cu caracter personal.

5.2. Studiu de caz privind utilizarea Learning Analytics pentru analiza comportamentului și tipologiei utilizatorilor de site-uri educaționale

Cu scopul de a evidenția utilizarea analiticii învățării, am realizat un studiu de caz pentru site-ul personal de pregătire la informatică, <https://www.modinfo.ro>. Întrebările de cercetare care stau la baza acestuia sunt:

- I1: Cine sunt, de unde provin și câți utilizatori are modinfo.ro?
- I2: Pot fi identificate modele comportamentale în interacțiunile utilizatorilor cu site-ul?
- I3: Care sunt dispozitivele de pe care este accesat site-ul?
- I4: Care sunt cele mai accesate pagini ale site-ului?
- I5: Care sunt canalele prin care utilizatorii accesează site-ul?

Pentru realizarea studiului și găsirea răspunsurilor la întrebările de cercetare fixate, am conectat Google Analytics cu instrumentul DataStudio care transformă datele în rapoarte informative și tablouri de bord. În maniera standard pot fi obținute trei tipuri de rapoarte referitoare la prezentarea generală a segmentelor de public (Fig. 5), prezentarea generală a achizițiilor (Fig. 6) și prezentarea generală a comportamentelor (Fig. 7).

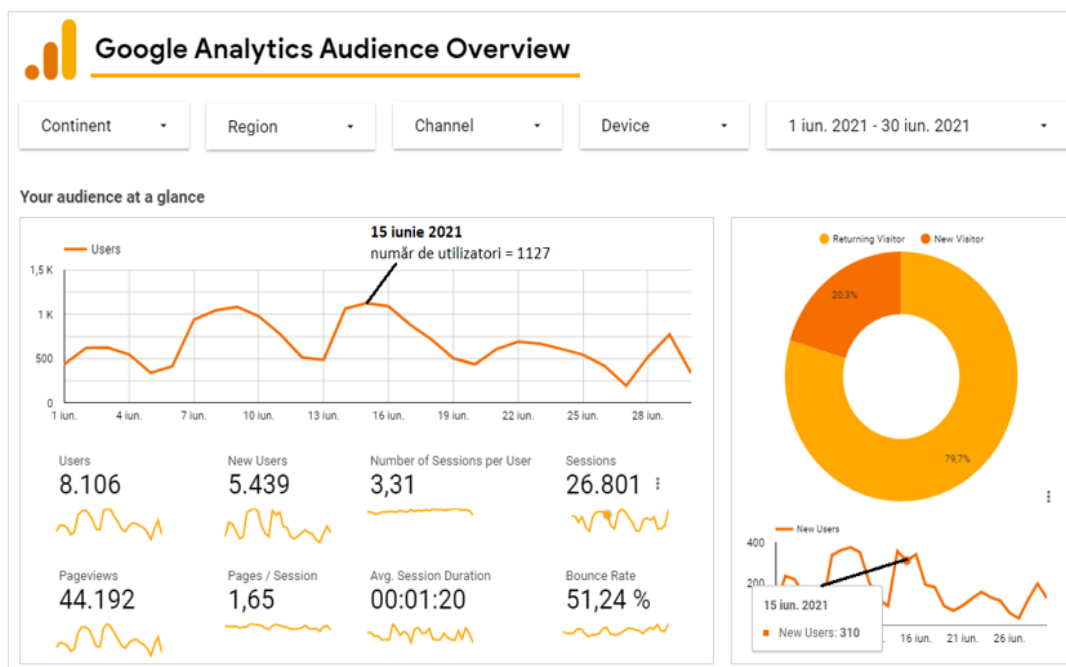


Figura 5. Prezentarea generală a segmentelor de public

Prezentarea generală a segmentelor de public din perioada aleasă pentru raportare (1 iunie 2021 – 30 iunie 2021) relevă că zona continentală cu cei mai mulți utilizatori ai site-ului este Europa de Est (26,7 K). Aceștia provin, în principal, din: București (4,5 K), Iași (3,1 K), Cluj (2,9 K),

Suceava (2 K), din totalul de 8106 din perioada selectată. Se observă că, pe 15 iunie 2021, există un maxim de utilizatori, 1127, dintre care 310 utilizatori noi (din 5439 raportați pentru luna iunie). Numărul total de pagini vizitate într-o sesiune este, în medie, de 1,65; în 15 iunie fiind 1,76. Rata de respingere este de 51,24% pentru întreaga perioadă. În 15 iunie se înregistrează o rată de respingere de 46,31%, fiind și cea mai mică.

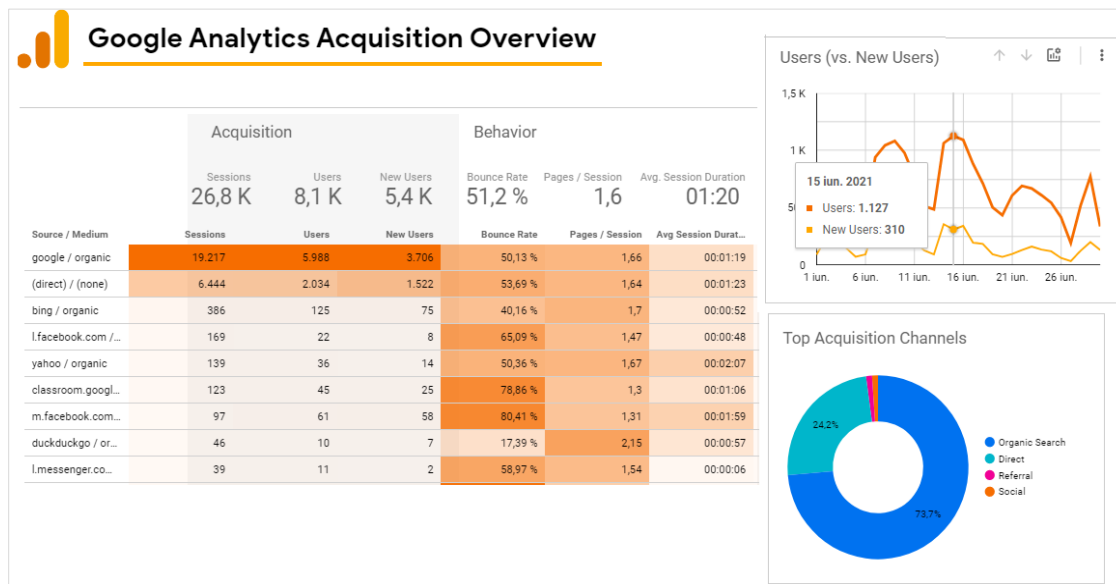


Figura 6. Prezentarea generală a achizițiilor

Prezentarea generală a achizițiilor (Fig. 6) arată că principalele canale prin care au fost efectuate achizițiile (accesările) sunt: căutarea organică pe Google (5988 de utilizatori), în proporție de 73,7% și accesarea directă a platformei (2034 utilizatori) în proporție de 24,2%. Rata de respingere este 50,13%, respectiv 53,69%, iar cea mai mică valoare a acestora este 80,41% și corespunde accesărilor efectuate prin intermediul rețelei Facebook, de pe un dispozitiv mobil.

Prezentarea generală a comportamentului utilizatorilor (Fig. 7) ilustrează faptul că cea mai accesată pagină a site-ului este *bac.php* ce conține subiecte și informații legate de acest examen. Au fost efectuate 28885 de vizualizări de pagină, reprezentând 65,4% din numărul total de vizualizări.

Un indicator foarte important despre comportamentul elevului pe site este rata de respingere. Ea este o măsură a procentului de utilizatori care au trimis o singură solicitare serverului și reprezintă raportul dintre numărul de sesiuni cu o singură pagină și numărul total de sesiuni. O valoare foarte mare, poate reprezenta o problemă, în cazul în care elevul trebuia să viziteze mai multe pagini, în sesiunea curentă. De exemplu, dacă trecerea la următoarele pagini este condiționată de promovarea unor itemi care atestă însușirea conținuturilor la nivel minimal, poate însemna că informațiile prezentate depășesc nivelul de pregătire, de înțelegere sau nu sunt interesante pentru elev.

Contribuțiile acestui studiu constau în: ilustrarea utilizării unui instrument gratuit pentru analitica datelor în context educațional și identificarea modalităților prin care, utilizând rapoartele privind prezentarea comportamentelor, a publicului și a achizițiilor, se poate obține o imagine a comportamentului elevilor în timpul folosirii unei aplicații de învățare. Valoarea predictivă a acestora constă în anticiparea unor acțiuni viitoare ale utilizatorilor pe baza rapoartelor generate pentru o perioadă anterioară de timp.

What do users see when they are in your website?

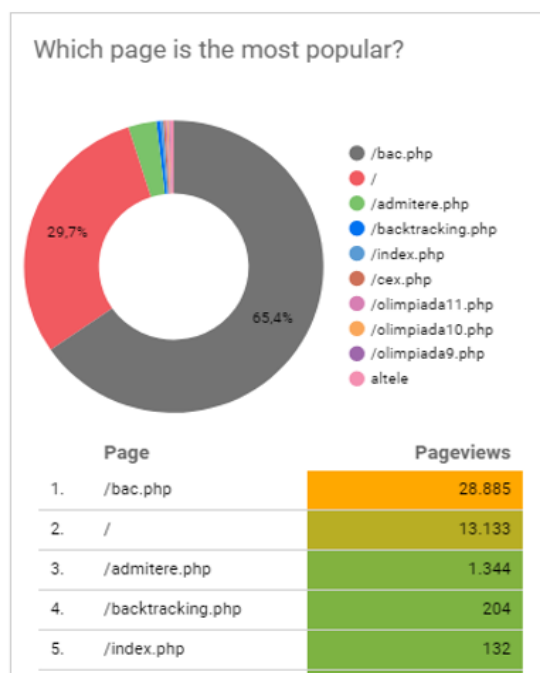


Figura 7. Prezentarea generală a comportamentelor

5.3. Investigație privind selecția tehnicilor de clasificare pentru predicția rezultatelor la Bacalaureat

5.3.1. Metodologie de lucru

Având la dispoziție rezultate la examenul de Bacalaureat din anii 2012-2019, colectate de la Colegiul Economic "Dimitrie Cantemir" din Suceava, a fost studiată performanța unor algoritmi pentru clasificarea elevilor în două clase: promovat/respins. Obiectivele cercetării sunt: identificarea celor mai bune metode de clasificare a datelor și găsirea unor posibile corelații între clasificarea rezultatelor și prezența sau nu în setul de date a anumitor caracteristici.

Pentru realizarea cercetărilor experimentale, am utilizat setul de date *EduBac* și instrumentul open source Orange, dezvoltat de cercetători de la Universitatea din Ljubliana pentru învățare automată și vizualizare a datelor de mari dimensiuni.

Procedura de lucru constă în crearea unui flux de date care va conduce către obținerea rezultatelor și vizualizare (Fig. 8).

Datele sunt introduse în fluxul operațional prin intermediul widget-ului *File* din fișierul Excel *bac.xlsx* care conține 2245 de instanțe și 13 caracteristici, așa cum au fost prezentate în capitolul 4.

Algoritmii de clasificare primesc date din fișierul de intrare, în urma unor operații de preprocesare care constau în selectarea instanțelor valide și stabilirea rolului și tipului caracteristicilor datelor. Tot în această fază este selectată și variabila țintă (*Target Variable*).

În scopul realizării validării încrucișate, a fost integrat în fluxul de lucru widget-ul *Test and Score*. Acesta are rolul de a calcula o serie de metrice de evaluare ale modelelor și de a le afișa sub formă tabelară.

Rezultatele evaluării sunt conectate la Matricea de confuzie (*Confusion Matrix*) care oferă 3 tipuri de vizualizare a rezultatelor: ca număr de instanțe, ca proporții ale predicțiilor, respectiv ale datelor reale.

În vederea obținerii de corelații între performanțele de clasificare ale modelelor utilizate și prezența sau nu a unor caracteristici în intrările din fluxul operațional, se vor repeta procedurile descrise mai sus pentru fiecare nou set de date obținut în urma procesărilor.

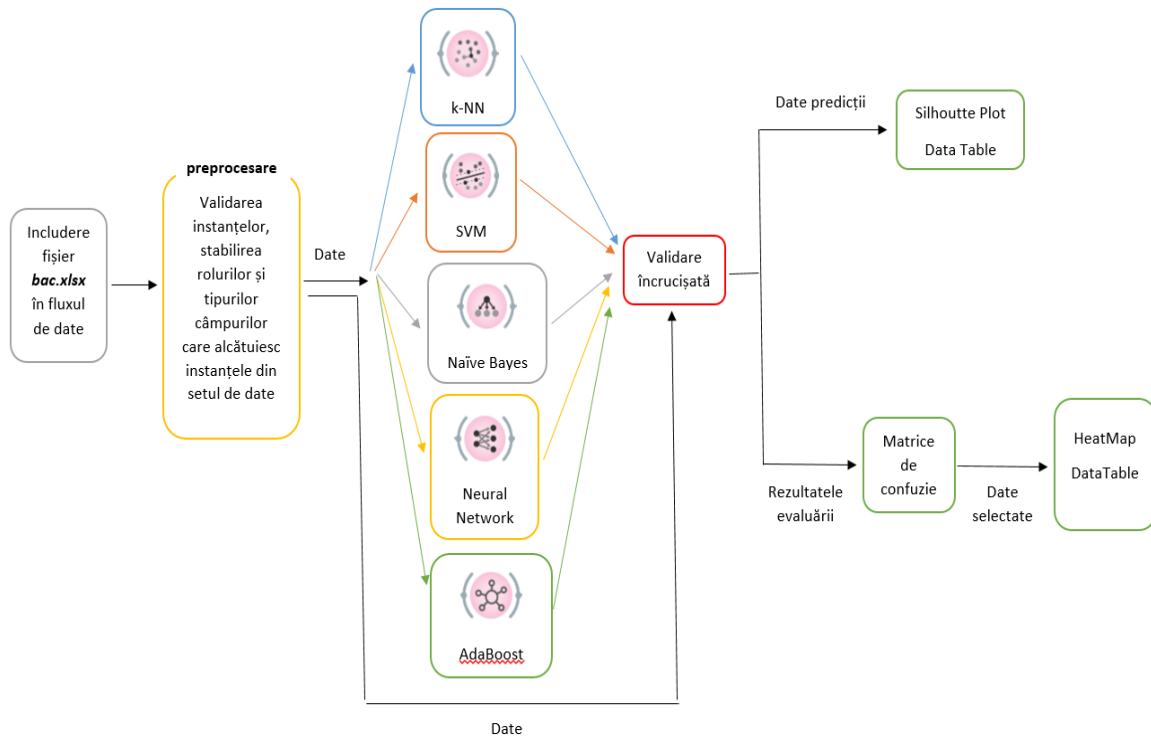


Figura 8. Fluxul de date

5.3.2. Rezultate obținute. Concluzii

Rezultatele obținute sunt foarte bune. În mod surprinzător, algoritmul AdaBoost dovedește performanțe incredibile de clasificare a rezultatelor evaluării elevilor la Bacalaureat, relevate de toate metricile utilizate. Următorul ca performanță este algoritmul Naive Bayes cu scorul de 0,999 pentru AUC. Rezultate apropiate obțin și algoritmi Neural Network și kNN cu scoruri de 0,998, respectiv 0,996 pentru AUC. Cele mai slabe rezultate le obține algoritmul SVM, cu o valoare de 0,987 pentru AUC (Fig. 9).

Model	AUC	CA	F1	Precision	Recall
kNN	0.996	0.984	0.984	0.985	0.984
SVM	0.987	0.957	0.956	0.958	0.957
Neural Network	0.998	0.982	0.982	0.982	0.982
Naive Bayes	0.999	0.986	0.986	0.987	0.986
AdaBoost	1.000	1.000	1.000	1.000	1.000

Figura 9. Rezultatele clasificării

Analizând matricile de confuzie asociate clasificării cu fiecare algoritm dintre cei precizați în tabelul 3 se pot face următoarele aprecieri:

- clasificarea corectă a promovaților se face în proporții cuprinse în intervalul [98,4%, 100%];
- clasificarea corectă a elevilor care nu au promovat examenul se face în proporții cuprinse în intervalul [94,7%, 100%];
- un procent de 0,7% dintre elevii respinși sunt clasificați drept promovați de către algoritmul Naïve Bayes, în timp ce 5,3% dintre elevii respinși sunt clasificați eronat de algoritmul kNN;
- un procent de 1,6% dintre elevii promovați sunt clasificați ca respinși cu algoritmul Naive Bayes;
- kNN clasifică greșit 0,3% dintre elevii promovați;
- Neural Network clasifică în mod eronat 4,1% dintre elevii respinși și 1,1% dintre cei promovați.

Tabel 3. Matricile de confuzie pentru algoritmi kNN, Neural Network, Naive Bayes, AdaBoost

kNN						AdaBoost					
			Predicted						Predicted		
			PROMOVAT RESPINS		Σ				PROMOVAT RESPINS		Σ
Actual	PROMOVAT		99.7 %	0.3 %	1680	Actual	PROMOVAT		100.0 %	0.0 %	1680
	RESPINS		5.3 %	94.7 %	565		RESPINS		0.0 %	100.0 %	565
	Σ		1705	540	2245		Σ		1680	565	2245
Neural Network						Naive Bayes					
			Predicted						Predicted		
			PROMOVAT RESPINS		Σ				PROMOVAT RESPINS		Σ
Actual	PROMOVAT		98.9 %	1.1 %	1680	Actual	PROMOVAT		98.4 %	1.6 %	1680
	RESPINS		4.1 %	95.9 %	565		RESPINS		0.7 %	99.3 %	565
	Σ		1685	560	2245		Σ		1657	588	2245

Un comportament diferit este înregistrat de algoritmul SVM (Fig. 10). El clasifică corect 99,5% dintre elevii promovați și 84,4% dintre cei respinși. 15,6% dintre elevii respinși sunt clasificați ca fiind promovați, în timp ce 0,5% dintre cei promovați sunt clasificați ca fiind respinși.

			Predicted		
			PROMOVAT RESPINS		Σ
Actual	PROMOVAT		99.5 %	0.5 %	1680
	RESPINS		15.6 %	84.4 %	565
	Σ		1760	485	2245

Figura 10. Matricea de confuzie pentru algoritmul SVM

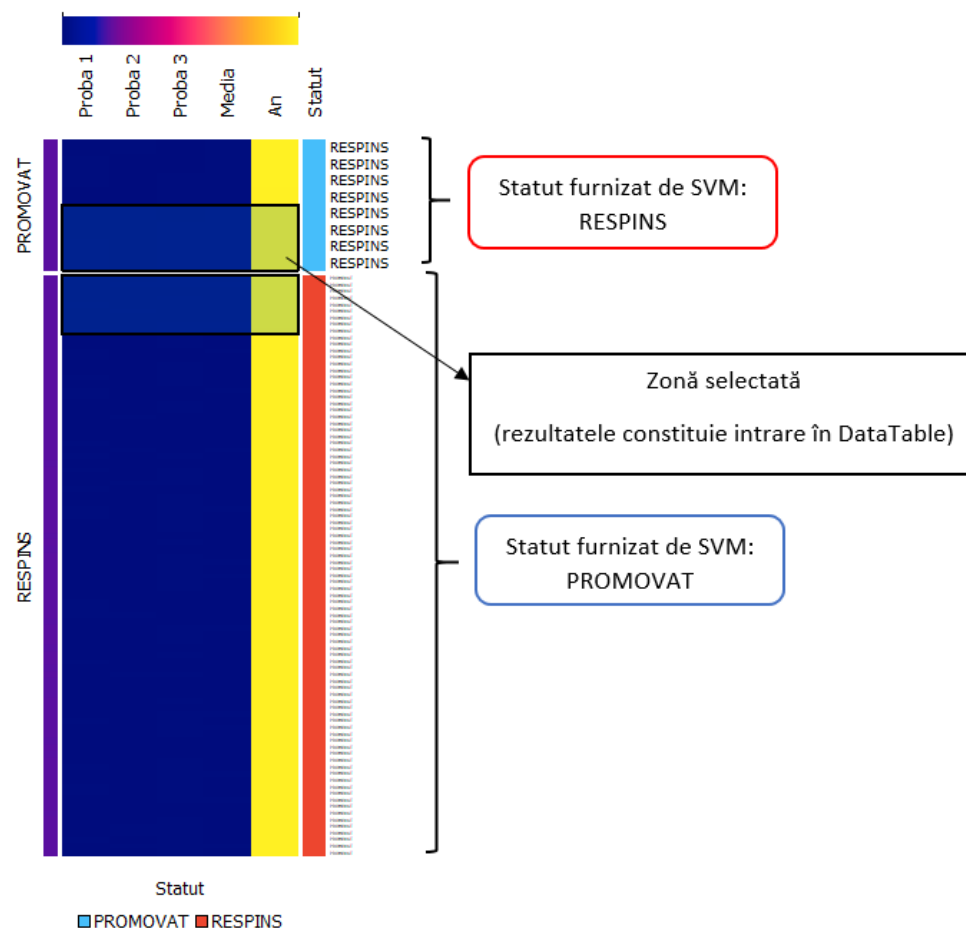


Figura 11. Vizualizarea selecției subsetului de date incorect clasificat în matricea de confuzie cu instrumentul HeatMap

	Statut	Statut(SVM)	Proba 1	Proba 2	Proba 3	Media	Profil	Proba D	Promotia curenta	Specializarea	Sex	Competențe lingvistice	Competențe digitale	An
1	PROMOVAT	RESPINS	5.40	7.40	5.45	6.08	Servicii	Logică, argumentare și comunicare	DA	Tehnician în activități economice	feminin	Utilizator Mediu	Utilizator Mediu	2017
2	PROMOVAT	RESPINS	5.60	5.55	7.80	6.31	Servicii	Logică, argumentare și comunicare	DA	Tehnician în gastronomie	masculin	Utilizator Avansat	Utilizator Mediu	2016
3	PROMOVAT	RESPINS	6.70	5.10	7.45	6.41	Servicii	Geografie	DA	Tehnician în activități de comerț	feminin	Utilizator Mediu	Utilizator Mediu	2014
4	PROMOVAT	RESPINS	6.40	5.10	6.85	6.11	Servicii	Geografie	NU	Tehnician în gastronomie	feminin	Utilizator Experimentat	Utilizator Experimentat	2014
5	RESPINS	PROMOVAT	5.40	5.00	7.10	5.83	Servicii	Geografie	DA	Tehnician în activități economice	masculin	Utilizator Experimentat	Utilizator Mediu	2019
6	RESPINS	PROMOVAT	5.00	6.55	6.40	5.98	Servicii	Logică, argumentare și comunicare	DA	Tehnician în activități economice	feminin	Utilizator Avansat	Utilizator Mediu	2019
7	RESPINS	PROMOVAT	5.00	5.00	5.75	5.25	Servicii	Logică, argumentare și comunicare	DA	Tehnician în activități de comerț	masculin	Utilizator Mediu	Utilizator Mediu	2019
8	RESPINS	PROMOVAT	5.05	5.00	6.60	5.55	Servicii	Logică, argumentare și comunicare	DA	Tehnician în turism	feminin	Utilizator Experimentat	Utilizator Mediu	2019
9	RESPINS	PROMOVAT	6.30	5.00	5.05	5.45	Servicii	Economie	DA	Tehnician în administrație	feminin	Utilizator Experimentat	Utilizator Avansat	2019
10	RESPINS	PROMOVAT	5.30	6.90	5.20	5.80	Real	Informatică MI C/C++	DA	Matematica-Informatica	masculin	Utilizator Mediu	Utilizator Mediu	2019
11	RESPINS	PROMOVAT	-2.00	-2.00	-2.00	0.00	Servicii	Logică, argumentare și comunicare	DA	Tehnician în gastronomie	masculin	Absent	Absent	2019
12	RESPINS	PROMOVAT	5.50	5.65	6.70	5.95	Servicii	Logică, argumentare și comunicare	DA	Tehnician în activități economice	masculin	Utilizator Experimentat	Utilizator Mediu	2019
13	RESPINS	PROMOVAT	6.20	5.10	6.35	5.88	Servicii	Geografie	DA	Tehnician în gastronomie	masculin	Utilizator Experimentat	Utilizator Avansat	2018

Figura 12. Vizualizarea rezultatelor selecției efectuate în HeatMap cu instrumentul DataTable

Acesta clasifică ca fiind promovat un elev care nu s-a prezentat la niciuna din probele examenului. Elevul a studiat la Profil *Servicii*, Specializarea *Tehnician în gastronomie*, în anul 2019. Vizualizarea rezultatelor a fost posibilă prin: conectarea instrumentului HeatMap la matricea de confuzie în care a fost selectat subsetul de date incorect clasificat de algoritmul SVM (Fig. 11) și conectarea la HeatMap a instrumentului Data Table (Fig. 12) pentru a obține detalii asupra datelor. Cu scopul de a analiza legătura dintre rezultate și detaliile despre datele elevilor, am folosit instrumentul Silhouette Plot (Fig. 13), pe care l-am conectat la intrare cu rezultatele clasificărilor și la ieșire cu instrumentul Data Table. Au fost alese spre exemplificare 6 instanțe: cea mai bine și cea mai slab clasificată instanță cu statut PROMOVAT, cea mai bine și cea mai slab clasificată instanță cu statut RESPINS și două instanțe aleatorii (Tabelul 4).

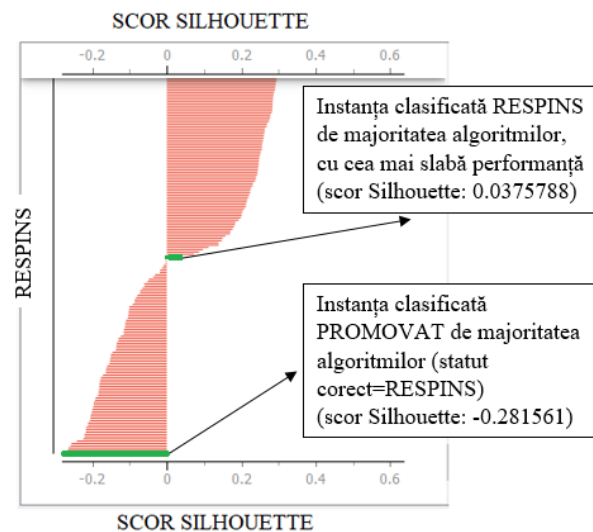


Figura 13. Vizualizarea rezultatelor cu Silhouette Plot. Evidențierea unor instanțe cu statut RESPINS.

Instanța 1 obține o clasificare corectă cu toți algoritmii, cu un scor de siluetă de 0.635321 și scoruri pentru clasa PROMOVAT de 1 (kNN, Neural Network și AdaBoost), respectiv 0.999817 (SVM) și 0.999999 (Naive Bayes). Se observă că notele la cele trei probe sunt foarte echilibrate (8.40, 8.10 și 8.75). Instanța descrie un elev de sex *feminin*, promoția curentă, de la specializarea *Tehnician în activități economice*, profil *Servicii*, anul susținerii examenului fiind 2016. Calificativele obținute la probele de *Competențe lingvistice* și *Competențe digitale* sunt: Utilizator Experimentat și Utilizator Avansat.

Tabel 4. Evidențierea rezultatelor obținute în clasificarea a 6 instanțe.

	Instanța 1 (cea mai bine clasificată PROMOVAT)	Instanța 2	Instanța 3 (cea mai bine clasificată RESPINS)	Instanța 4	Instanța 5 (cea mai slab clasificată RESPINS)	Instanța 6 (cea mai slab clasificată PROMOVAT)
Statut:	PROMOVAT	PROMOVAT	RESPINS	RESPINS	RESPINS	PROMOVAT
kNN	PROMOVAT	PROMOVAT	RESPINS	RESPINS	PROMOVAT	PROMOVAT
SVM	PROMOVAT	PROMOVAT	RESPINS	PROMOVAT	PROMOVAT	PROMOVAT
Naive Bayes	PROMOVAT	RESPINS	RESPINS	RESPINS	RESPINS	RESPINS
Neural Network	PROMOVAT	RESPINS	RESPINS	PROMOVAT	PROMOVAT	RESPINS

	Instanța 1 (cea mai bine clasificată PROMOVAT)	Instanța 2	Instanța 3 (cea mai bine clasificată RESPINS)	Instanța 4	Instanța 5 (cea mai slab clasificată RESPINS)	Instanța 6 (cea mai slab clasificată PROMOVAT)
AdaBoost	PROMOVAT	PROMOVAT	RESPINS	RESPINS	RESPINS	PROMOVAT
kNN (PROMOVAT)	1	0.6	0	0.2	0.8	0.6
kNN (RESPINS)	0	0.4	1	0.8	0.2	0.4
SVM (PROMOVAT)	0.999817	0.284197	1.95981e-08	0.25692	0.913704	0.284197
SVM (RESPINS)	0.00018271	0.715803	1	0.65344	0.0862962	0.715803
Naive Bayes (PROMOVAT)	0.999999	0.0059232	1.04058e-06	1.42353e-06	0.00691154	0.0059232
Naive Bayes (RESPINS)	9.79227e-07	0.994077	0.999999	0.999999	0.993088	0.994077
Neural Network (PROMOVAT)	1	0.39313	1.28786e-14	0.579533	0.697232	0.39313
Neural Network (RESPINS)	2.52205e-10	0.60687	1	0.420467	0.302768	0.60687
AdaBoost (PROMOVAT)	1	1	2.22045e-16	2.22045e-16	2.22045e-16	1
AdaBoost (RESPINS)	2.22045e-16	2.22045e-16	1	1	1	2.22045e-16
Silhouette (Statut)	0.635321	0.153388	0.495545	-0.102812	-0.281561	0.153388
Proba 1	8.40	5.00	5.10	6.15	5.80	5.00
Proba 2	8.10	5.00	1.45	5.00	6.80	5.00
Proba 3	8.75	8.00	5.00	6.00	5.05	8.00
Media	8.41	6.00	0.00	5.71	5.88	6.00
Profil	Servicii	Servicii	Servicii	Servicii	Real	Servicii
Proba D	Geografie	Geografie	Geografie	Geografie	Informatică	Geografie
Promoția curentă	DA	NU	DA	DA	DA	NU
Specializarea	Tehnician în activități economice	Tehnician în activități de comerț	Tehnician in activitati economice	Tehnician în activități de comerț	Matematica- Informatica	Tehnician in activitati de comert
Sex	feminin	feminin	feminin	feminin	masculin	feminin
Competențe lingvistice	Utilizator Experimentat	Utilizator Mediu	Utilizator Mediu	Utilizator Mediu	Utilizator Experimentat	Utilizator Mediu
Competențe digitale	Utilizator Avansat	Utilizator Avansat	Utilizator Avansat	Utilizator Avansat	Utilizator Experimentat	Utilizator Avansat
An	2016	2013	2013	2013	2017	2013

În aceeași manieră, s-au făcut corelații și pentru celelalte 5 instanțe analizate. Concluzia, fieească de altfel, este că algoritmi clasifica cu mai multă ușurință elevii cu rezultate mai mici decât 5 și mai mari decât 6.50.

Prin execuția algoritmilor pentru subseturile de date alcătuite doar din anul 2019, respectiv anii 2018 și 2019, am constatat următoarele: AdaBoost obține aceleași performanțe extraordinare, SVM obține scoruri mai mari, anume 0,993 față de 0,987 pentru setul complet de date. kNN obține performanțe mai bune pentru clasificarea elevilor din anul 2019, metrica AUC având scorul 0,999. În cazul algoritmului Naive Bayes am identificat o scădere a performanțelor de clasificare pentru anii 2018 și 2019, la 0,995 – 0,996 față de 0,999 pentru toți anii (Tabelul 5).

Tabel 5. Rezultate comparative privind performanța clasificării pentru dimensiuni diferite ale setului de date

Algoritm de clasificare	Scor pentru setul de date ce cuprinde toți anii					Scor pentru setul de date ce cuprinde anul 2019					Scor pentru setul de date ce cuprinde anii 2018 și 2019				
	AUC	CA	F1	P	R	AUC	CA	F1	P	R	AUC	CA	F1	P	R
kNN	0.996	0.984	0.984	0.985	0.984	0.999	0.972	0.972	0.973	0.972	0.995	0.979	0.979	0.979	0.979
SVM	0.987	0.957	0.956	0.958	0.957	0.993	0.966	0.966	0.966	0.966	0.993	0.954	0.953	0.954	0.954
Neural Network	0.998	0.982	0.982	0.982	0.982	0.996	0.969	0.969	0.969	0.969	0.995	0.978	0.978	0.978	0.978
Naive Bayes	0.999	0.986	0.986	0.987	0.986	0.982	0.954	0.955	0.958	0.954	0.997	0.982	0.983	0.983	0.982
AdaBoost	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Au fost create și alte subseturi de date: fără *Competențe digitale*, fără *Competențe lingvistice*, fără *Sex* și un set în care sunt prezente doar aceste caracteristici. Rezultatele obținute, după rularea algoritmilor, au relevat următoarele: există slabe corelații între rezultatele la competențele digitale sau lingvistice, genul candidatului și statutul final al acestuia. Excepție face, din nou, SVM care clasifică mai bine în absența caracteristicii *Sex* din datele de intrare. În absența altor informații privind candidații, în afara calificativelor obținute la competențe și genul candidatului, au fost obținute rezultatele: SVM are scorul cel mai mic de clasificare – 0,473, AdaBoost are performanțe cu mult reduse – 0,651, dar foarte apropiate de cele obținute de ceilalți algoritmi. Scorurile acestora sunt: kNN – 0,605, NN – 0,669 și NB – 0,623. Date fiind acestea, se poate aprecia că un elev va fi promovat sau respins numai pe baza studierii situației școlare la limbi moderne și Tehnologia Informației și a Comunicațiilor cu o probabilitate de 60%.

Contribuțiile acestei cercetări constau în: analiza comparativă a performanțelor de clasificare a 5 algoritmi pentru a stabili oportunitatea utilizării unora dintre ei, identificarea unor corelații între rezultate și caracteristicile datelor și formularea unor concluzii privind acțiuni și direcții viitoare pe baza predicțiilor efectuate.

Dat fiind comportamentul bizar al algoritmului SVM, se pot face studii viitoare care să analizeze motivele rezultatelor clasificărilor cu acesta.

Ușoarele abateri de la precizia de clasificare obținute pentru subseturile ce conțin doar anii 2018 – 2019 și 2019 semnifică faptul că nu există modificări majore privind profilul elevului în perioada analizată. Stabilirea și aplicarea unor proceduri de creștere a rezultatelor școlare precum: planuri remediale individualizate, implicarea părinților, ore suplimentare la anumite materii, etc ar putea avea succes pentru o perioadă de cel puțin 10 ani.

O analiză mai profundă a corelațiilor cu rezultatul final obținut de elevi se poate efectua luând în calcul și evoluția școlară a acestora, la toate materiile, de-a lungul celor 4 ani de liceu, situația familială, cea materială, implicarea în activitățile școlii, profesorii de la clasă și alți factori. Aceasta poate constitui o direcție viitoare de cercetare. Identificarea cu precizie a profilului elevului de 6, a celui de 7, etc poate facilita elaborarea unor măsuri eficiente pentru realizarea progresului școlar.

5.4. Implementarea proiectelor de analiză a sentimentelor folosind serviciul AutoML Language din Google Cloud

Deoarece construirea modelelor de învățare automată este un proces complex și dificil care necesită abilități științifice avansate, în scopul identificării unor soluții mai facile, mai ușor de pus în practică, în sistemul preuniversitar, chiar și de către acei profesori care nu sunt specialiști în domeniu, am dorit să examinez și să prezint potențialul de utilizare al AutoML (Automated Machine Learning) pentru a clasifica opiniile elevilor despre școală. Din oferta de instrumente disponibilă în prezent, am ales serviciul AutoML Language. Din păcate, acesta nu este suficient de transparent privind arhitectura modelului și reglarea hiperparametrilor, însă poate constitui o direcție ulterioară de cercetare crearea unui model propriu, ale cărei performanțe să poată fi studiate comparativ pe seturi de date de dimensiuni mici, dar și pe volume mari de date, unde AutoML pare a fi performante.

Folosind Google Cloud Storage, a fost creat bucket-ul *sentiment-analysis-1* unde vor fi încărcate datele. Acesta folosește o locație de tip *Multi-region*, tipul de criptare *Google-managed key* și conține 3 foldere: *Date-opinii*, *Results_Sentiment-Analysis*, *Sentiment-Analysis* (Fig. 14).

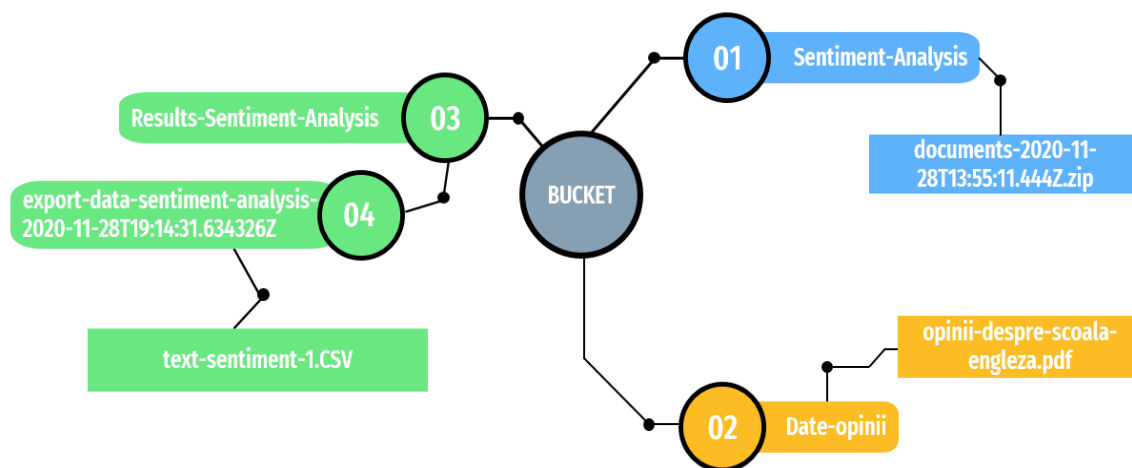


Figura 14. Conținut bucket *sentiment-analysis-1*

5.4.1. Modelul ASG

Modelul ASG (Analiza Sentimentelor cu Google) a fost numit astfel în corelație cu scopul utilizării lui și platforma folosită. Pentru antrenarea, validarea, evaluarea acestuia am folosit setul de date *claritin_october_twitter_side_effects*, prezentat în capitolul 4. Modelul a fost utilizat pentru etichetarea setului de date *191-opinii-despre-școală*, de asemenea prezentat anterior.

Performanțele obținute în clasificarea textelor sunt destul de slabe; Precision=Recall=64,79% la nivelul tuturor claselor de sentiment.

Cel mai bine sunt clasificate pozitiv sentimentele cu scorul 0 (82%), iar cel mai slab sunt clasificate pozitiv sentimentele care au scorul 4 (32%). 6% dintre fișierele cu scor de sentiment 0, dintre cele utilizate la evaluare, sunt clasificate eronat cu scorul de sentiment 1, iar 12% cu scorul de sentiment 2. Performanțele cele mai slabe de clasificare par să fie în legătură cu datele etichetate cu un scor de sentiment 4, 63% din documente sunt asociate eronat cu scorul de sentiment 3, urmate de fișierele etichetate cu scorul 2 care sunt mapate eronat unui scor de sentiment 3 în proporție de 41% (Fig. 15).

Actual score	Predicted score	Sentiment score 0	Sentiment score 1	Sentiment score 2	Sentiment score 3	Sentiment score 4
Sentiment score 0	82%	6%	12%	-	-	-
Sentiment score 1	2%	45%	30%	23%	-	-
Sentiment score 2	-	5%	54%	41%	-	-
Sentiment score 3	-	3%	16%	78%	2%	-
Sentiment score 4	-	5%	-	63%	32%	-

Figura 15. Matricea de confuzie pentru toate scorurile de sentiment

sentiment_analysis_table				Distribuți		?		⚙	
Date de conectare la date: Proprietar				Actualitatea datelor: 12 ore		Accesul vizualizărilor comunității: Activat		Editarea câmpurilor din rapoarte: Activată	
MODIFICĂȚI CONEXIUNEA				FILTRAȚI ÎN FUNCȚIE DE ADRESA DE E-MAIL				ADĂUGAȚI UN CÂMP	
ADĂUGAȚI UN PARAMETRU									
Câmp	Tip	Cumulare prestabilită	Descriere						
PARAMETRI (2)									
Content	RBC Text		Niciunul						
Predicted_score	123 Număr		Sumă						
VALORI (1)									
Record Count	123 Număr		Automat						

Figura 16. Crearea unui raport folosind Data Studio conectat la produsul BigQuery din Google Cloud

Un alt set de date, pe care l-am denumit *191-opinii-despre-școală*, conține 191 de opinii pe care le-am colectat prin intermediul unui chestionar de la elevi de clasa a 11-a și a 12-a de la specializarea matematică-informatică, din Suceava. Acesta este neetichetat. Fiecare opinie a fost asociată unui scor de sentiment prin utilizarea experienței modelului ASG. Pentru vizualizarea rezultatelor într-o manieră mai prietenoasă, a fost conectat instrumentul Data Studio la serviciul BigQuery din Google Cloud de unde sunt preluate structura tabelului și datele memorate în bucket-ul proiectului (Fig. 16).

Analizând raportul pe care l-am construit pe baza rezultatelor, am constatat că 77 din cele 191 de texte utilizate sunt asociate cu clasa de sentiment 3. Acestea corespund unor opinii ușor pozitive ale elevilor în legătură cu școala de tipul: *"It does not make me happy, but I Like it"*. Întrucât precizia de clasificare a modelului este de aproape 80% pentru textele etichetate cu scorul 3, pot aprecia că aproximativ 40% dintre elevi au opinii relativ bune în privința școlii.

Cea mai bună precizie de clasificare este asociată cu scorul 0. Doar 5 elevi au exprimat opinii atât de negative. Ei reprezintă un procent de doar 2,6%, iar opiniile sunt de tipul: *"I hate school"*.

72 dintre elevii chestionați au păreri relativ proaste privind școala, etichetate de model cu scorul 1. Având în vedere slaba performanță obținută de model în clasificarea acestui tip de texte (45%), nu este de mirare că opinii precum *"I am glad to see with my colleagues and participate in some hours but I hate other hours because of the teachers"* și *"An ok environment, relatively quiet. It creates an indifferent state to me"* sunt asociate cu scorul 1, deși consider că ar trebui să aibă mai curând scorul 2.

12 opinii sunt asociate, prin utilizarea modelului, cu clasa 4. Un exemplu corect atribuit este: *"It creates me a feeling of happiness"*. Având în vedere că pentru această clasă, performanța modelului este cea mai mică, 32%, rezultatele cu greu pot fi luate în considerare. Un exemplu în acest sens sunt două opinii etichetate cu 4, deși apreciez că exprimă sentimente din clasele 2-3 și 1: *"Essential necessity, even if sometimes unpleasant"* și *"It's not my favorite place"*.

Cu ajutorul rapoartelor de vizualizare se pot extrage mai ușor informații privind rezultatele obținute. Concluzia acestui studiu este că aproximativ 47% dintre elevi au opinii bune și foarte bune despre școală, 11,5% sunt indiferenți și 40,31% detestă școala, cu o precizie de clasificare de aproximativ 65%.

5.4.2. Modelul Opinii

La antrenarea, validarea și evaluarea modelului ASG au fost utilizate date deja etichetate, disponibile gratuit, dar care nu provin din domeniul educațional, deci neavând același specific. O ipoteză naturală este aceea că rezultatele obținute de un model care folosește aceeași arhitectură, ar putea fi cu mult îmbunătățite dacă datele utilizate pentru antrenare sunt similare cu acelea din setul ce urmează a fi etichetat pe baza experienței modelului. Pornind de la acest deziderat, am realizat un chestionar complex pe care l-am distribuit la nivel național. Scopul cercetării este acela de a obține cunoștințe credibile, valoroase privind opiniile exprimate de elevi care învață la licee cu diferite profiluri și specializări din România.

Modelul Opinii, denumit astfel după setul de date utilizat pentru antrenament și evaluare, obține rezultate mult mai bune în clasificarea textelor decât ASG, având aceeași structură cu acesta; Precision=Recall=86,51%.

Actual score	Predicted score	Sentiment score 0	Sentiment score 1	Sentiment score 2	Sentiment score 3	Sentiment score 4
Sentiment score 0	87%	-	-	13%	-	-
Sentiment score 1	-	93%	7%	-	-	-
Sentiment score 2	-	8%	80%	4%	8%	-
Sentiment score 3	-	-	9%	77%	14%	-
Sentiment score 4	-	-	4%	4%	92%	-

Figura 17. Matricea de confuzie pentru modelul Opinii

Opiniile asociate scorului de sentiment 1 sunt etichetate corect în proporție de 93%. Numai 77% dintre datele folosite pentru evaluare sunt etichetate corect cu scorul 3. 14% din aceste opinii sunt atribuite de către model, în mod eronat, clasei de sentiment cu scorul 4. 13% dintre opiniile care, în mod real au fost etichetate cu un scor de sentiment 0, sunt atribuite de model clasei cu scorul de sentiment 3.

Se observă că, opiniile cu scorurile 0 și 1, sunt asociate eronat câte unei singure clase de sentiment; 7% din datele etichetate cu scorul 1 sunt asociate clasei de sentiment 2. Sentimentele etichetate corect cu scorul 2 reprezintă 80% din datele utilizate pentru evaluarea modelului. Restul de 20% din aceste opinii sunt asociate de model sentimentelor cu scorurile 1 (8%), 3 (4%) și 4 (8%) (Fig. 17).

Prin utilizarea modelului pe setul de date *191-opinii-despre-școală*, au fost obținute rezultatele: 125 opinii au fost etichetate de model cu scorul de sentiment 0, 6 cu scorul 1, 52 cu scorul 3, 7 cu scorul 4 și doar un document cu scorul 2

5.4.3. Comparație între modelele ASG și Opinii. Concluzii

Spre deosebire de modelul antrenat pe date care nu sunt din educație și care a obținut o precizie de clasificare de aproximativ 65%, modelul antrenat pe setul propriu de date, provenit de la elevii români are o precizie de 86,51%. Rezultă imediat două concluzii: este important setul de date folosit la antrenarea modelului și informațiile relevate de analiza sentimentelor cu modelul Opinii vor fi mult mai credibile.

La nivelul fiecărei clase de sentiment, am putut observa următoarele:

- ambele modele obțin performanțe asemănătoare de clasificare a documentelor etichetate cu scorul 0 și 3. Pentru scorul 0, modelul ASG obține 82%, iar Opinii 87%. Pentru scorul 3, modelul ASG obține 78%, în timp ce Opinii 77%.
- diferențe semnificative se observă la rezultatele clasificării pentru clasa de sentimente 4 – modelul ASG obține 32%, iar Opinii 92%.
- pentru clasele 1 și 2 am obținut rezultatele: modelul ASG are 45%, respectiv 54% precizie de clasificare. Modelul Opinii obține 93%, respectiv 80%.

La fel ca ASG și modelul Opinii asociază textul *"It does not make me happy, but I Like it"* cu scorul de sentiment 3.

Opinia *"I am glad to see with my colleagues and participate in some hours but I hate other hours because of the teachers"*, etichetată cu 1 de ASG a fost asociată de Opinii scorului de sentiment 3. Este interesant rezultatul întrucât eu aș fi etichetat cu scorul 2 acest text. Și mai interesantă este asocierea opiniei *"An ok environment, relatively quiet. It creates an indifferent state to me"* cu scorul 0 de către Opinii și cu scorul 1 de ASG, în timp ce eu aș fi etichetat cu scorul de sentiment 2 și această opinie.

Un alt rezultat interesant este asocierea opiniei *"It creates me a feeling of happiness"* cu scorul 3 de către Opinii și cu 4 de către ASG. Aș asocia cu scorul 4.

Modelul Opinii asociază cu sentimentul cel mai negativ despre școală opinia *"Essential necessity, even if sometimes unpleasant"*, în timp ce ASG o atribuie clasei 4. Aș asocia cu 2 sau 3.

"It's not my favorite place" este clasificată ca fiind cu scorul 0 de către Opinii; aș eticheta cu 1.

Modelul Opinii clasifică 65,44% din cele 191 opinii ca fiind profund negative. Doar 30,89% dintre textele analizate sunt asociate cu sentimente bune și foarte bune în legătură cu școala. Modelul ASG atribuie un scor 2 la nivelul tuturor opiniilor, în timp ce Opinii atribuie scorul 1.

Contribuțiile acestor cercetări constau în: utilizarea AutoML Language și a altor servicii din Google Cloud pentru realizarea a două proiecte de analiza sentimentelor din date de tip text, și formarea, cu o precizie destul de bună a unei imagini privind realitatea sentimentelor pe care le au elevii români despre școală. Acest gen de experiment poate fi utilizat pentru a explora mesajele pe care elevii le scriu pe o platformă educațională, fie ca și comentarii la activități de instruire, fie sub formă de chat sau forumuri de discuții. Cunoștințele obținute pot genera acțiuni care să îmbunătățească calitatea învățării.

5.5. Analiza sentimentelor cu modelele ASO - Ekman și ASO – Plutchik

5.5.1. Emoțiile. Modelele Ekman și Plutchik

Paul Ekman consideră că fiecare emoție poate fi asociată cu o mimică facială specifică. Deoarece membrii aceleiași culturi își exprimă similar emoțiile, există mai puține dificultăți în comunicarea dintre aceștia, față de transmiterea emoțiilor între componenții societăților diferite. Emoția este rezultat al evoluției. Manifestările non-verbale ale acestora sunt clasificate în raport cu 6 emoții de bază: *fericire (happiness)*, *tristețe (sadness)*, *surpriză (surprise)*, *dezgust (disgust)*, *furie (anger)*, *frică (fear)* [318].

Robert Plutchik consideră că emoțiile pot fi puse în perechi antagonice precum bucurie-tristețe. În clasificarea acestora are în vedere următoarele postulate: emoțiile sunt specifice atât oamenilor cât și animalelor, au o evoluție istorică și au jucat de-a lungul timpului un rol foarte important în adaptarea la mediu și supraviețuirea speciilor. El consideră de asemenea că aceeași emoție se poate manifesta cu un grad diferit de intensitate. Plutchik propune un model structural, cunoscut drept roata emoțiilor. El identifică 8 emoții de bază: *bucurie (joy)*, *tristețe (sadness)*, *frică (fear)*, *încredere (trust)*, *furie (anger)*, *dezgust (disgust)*, *surpriză (surprise)* și *anticipare (anticipation)* [236].

5.5.2. Modelele ASO-Ekman și ASO-Plutchik

Pentru efectuarea cercetărilor experimentale, am utilizat setul de date *191-opinii-despre-școală*, prezentat în capitolul 4. Opiniile elevilor au fost traduse în engleză și salvate în fișierul Excel *opinii-despre-scoala-engleza.xlsx*. O alternativă la formatul comma-separated values (.CSV), larg acceptată pentru schimbul de date dintre platforme sau între baze de date și fișiere în format tabelar, este tab-separated values (.TSV sau .TAB). În acest format de fișier, fiecare rând corespunde unei înregistrări din tabelul Excel și fiecare câmp de pe un rând corespunde unei coloane și este separat prin caracterul TAB de celelalte de pe același rând.

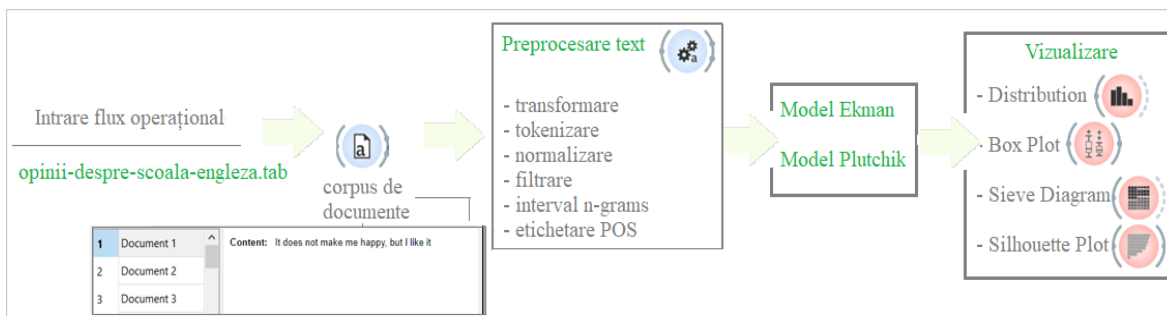


Figura 18. Fluxul operațional

Utilizând înregistrările din Excel, a rezultat fișierul *opinii-despre-scoala-engleza.tab*. Datele au fost incluse într-un flux operațional și transformate într-un corpus de documente, în acest fel fiecare rând constituind un document separat identificabil prin ID-ul de forma *Document nr.* În continuare, am conectat widget-ul *Tweet Profiler*. Acesta preia informațiile și le trimite către un server unde modelul Ekman sau Plutchik va calcula probabilitățile sau scorurile și va asocia fiecărui document emoția dominantă. Am utilizat modelele Ekman (Multi Class) și Plutchik (Multi Class). Pentru vizualizarea rezultatelor, am inserat în fluxul de lucru câteva instrumente de vizualizare: *Distributions, BoxPlot, Sieve Diagram, Silhouette Plot* (Fig. 18).

58,6% dintre documente au fost clasificate de modelul Ekman ca fiind *Joy*.

Cu modelul Plutchik un procent de 68% dintre documente au fost clasificate unor emoții pozitive, respectiv 37,6% *Joy* și 32,4% *Trust*.

În general, modelul Ekman asociază toate opiniile ce exprimă indiferența față de școală cu emoția *Anger*. Documentul 142 conține textul: *"I do not feel it helps me in my training"*. Este interesant faptul că, modelul ASG atribuie scorul de sentiment 1 acestui mesaj, iar modelul Opinii scorul 0. Ambele au fost prezentate în secțiunea anterioară. Cu toate diferențele arhitecturale, modelele detectează în egală măsură emoțiile negative.

Prin comparație, Plutchik asociază cu această emoție doar documentul 130: *"Just total hatred"*. Și acest text este clasificat similar de ASG, care-i atribuie scorul 1 și Opinii care-l asociază cu clasa de sentiment 0, deci profund negativ.

Observând diferența destul de importantă în ceea ce privește numărul de opinii asociate cu sentimente pozitive, prin comparație cu rezultatele obținute în Google Cloud, am analizat cu atenție documentele din corpus corespunzătoare emoției *Joy*. Am constatat că, atât Ekman cât și modelul Plutchik asociază cu emoția *Joy* opinii care, în mod evident, exprimă un sentiment negativ legat de școală.

Analizând textele clasificate, ca rezultat al utilizării fiecăruia dintre modelele descrise în lucrare, se constată că, cu modelul Opinii, 59 dintre opiniile elevilor, reprezentând doar 30,89% din cele 191, sunt pozitive în privința școlii. Spre deosebire, modelul Ekman asociază aproximativ 59% dintre opinii cu o emoție ce exprimă bucurie, ceea ce este departe de a fi adevărat.

Contribuția acestui studiu experimental constă în utilizarea instrumentului gratuit Orange pentru a clasifica opiniile elevilor în funcție de emoțiile de bază definite de psihologii Ekman și Plutchik. Au putut fi realizate analize comparative pe baza cărora am constatat că abordarea bazată pe lexicon, folosită în acest experiment obține rezultate mult mai slabe decât cea bazată pe învățarea automată, folosită pentru realizarea proiectului de analiză a sentimentelor cu modelul Opinii din Google Cloud.

5.6. Proiectarea și implementarea proceselor de analiză a sentimentelor din imagini

5.6.1. Analiza sentimentelor folosind AutoML Vision din Google Cloud

În cercetările efectuate, am utilizat serviciul AutoML Vision din Google Cloud pentru a crea două proiecte de analiză a sentimentelor din imagini. Modelele utilizate în cadrul acestor proiecte sunt antrenate și testate pe date proprii. Același set de imagini va fi etichetat diferit cu scopul de a analiza comparativ performanța celor două modele. Pentru modelul ASI – G4 (Analiza Sentimentelor din Imagini – Google 4) voi eticheta 4 emoții, iar pentru ASI - G2 (Analiza Sentimentelor din Imagini – Google 2) voi folosi două clase pentru a descrie emoțiile. Setul de date

utilizat este *Imagini-generale*. În figura 19 sunt descrise principalele etape parcurse în construirea, antrenarea și testarea modelului cu scopul identificării emoțiilor în datele etichetate.



Figura 19. Principalele etape pentru construirea modelului folosind AutoML Vision

Am utilizat tipul de criptare *Google-managed-key*. Ca metodă de clasificare am folosit *Single-Label Classification*. Am utilizat 1 nod cu buget de 16 ore pentru antrenarea modelului și 1 nod pentru implementarea acestuia. Imaginile au fost etichetate manual.

5.6.1.1. Modelul ASI - G4

Denumirea modelului pe care l-am utilizat a fost aleasă în corelație cu scopul propus și anume detectarea emoțiilor: fericire, încredere, teamă, tristețe.

În figura 20 se pot observa modificările Precision și Recall odată cu creșterea valorii pragului de încredere, de la valoarea de 0,5, când Precision și Recall sunt 56,25%, la 0,998 când Precision este 77,78% și Recall are valoarea 43,75%, măsurători efectuate la nivelul tuturor categoriilor etichetate.

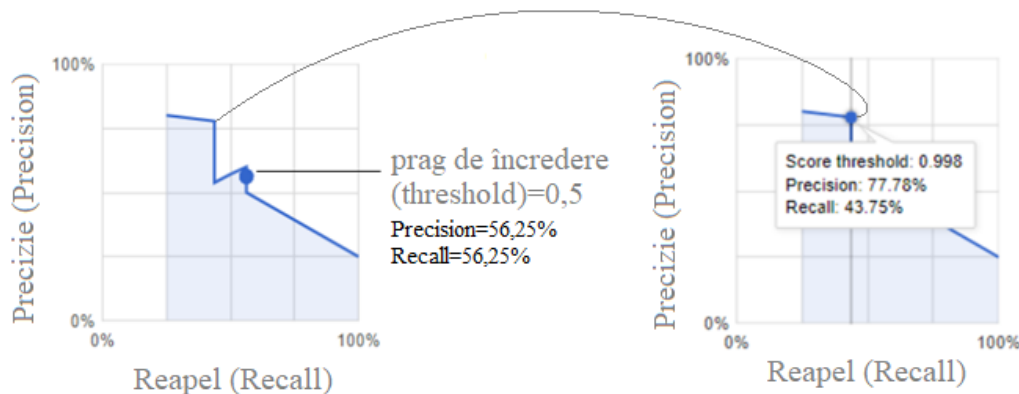


Figura 20. Curba Precision-Recall – evidențiere pentru valori ale pragului de încredere de 0,5 și 0,998



Figura 21. Modificarea Precision și Recall odată cu valorile pragului de încredere

În graficul din figura 21 se poate observa că valorile Precision și Recall rămân setate la 56,25% la variații destul de mari ale pragului de încredere, ceea ce înseamnă că modelul este stabil în legătură cu clasificările efectuate odată cu modificarea numărului de cazuri true pozitive și a celor false negative.

5.6.1.2. Modelul ASI – G2

Am utilizat două clase în locul celor patru și am încercat maparea imaginilor la una dintre caracteristici: *pozitiv*, *negativ*.

O primă observație privind rezultatele obținute este că precizia medie, Recall și Precision au valori semnificativ mai mari decât cele obținute cu modelul anterior, valori calculate la un prag de încredere de 0,5 și anume: PrecisionRecall=87,5%, Average precision=0,929.

Analizând curba de evoluție a Precision și Recall, se constată că, similar cu modelul ASI-G4, odată cu modificarea pragului de încredere, Precision și Recall tind să-și păstreze constante valorile deci clasificarea efectuată este stabilă la modificarea numărului de cazuri true pozitive și false negative (Fig. 22).

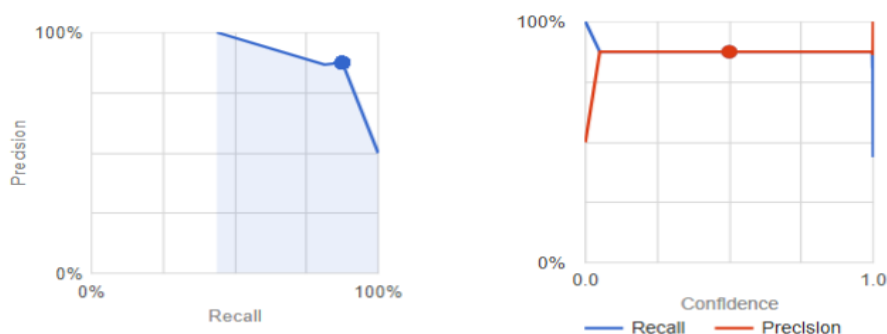


Figura 22. Curba Precision-Recall și curba evoluției Precision și Recall cu modificarea pragului de încredere, la nivelul celor două caracteristici: pozitiv, negativ

Identificarea unor concepte abstracte precum fericire și tristețe este mai dificil de efectuat decât împărțirea în două categorii a setului de date. Utilizarea unui model de clasificare a imaginilor într-o aplicație educațională ar avea mai mult succes în evidențierea unei stări emoționale pozitive sau negative sau în clasificarea folosind alte categorii precum: floare, natură, casă, etc.

Contribuțiile acestor cercetări experimentale constau în: utilizarea serviciilor din Google Cloud pentru analiza sentimentelor din imagini, realizarea unei analize comparative între două modele și identificarea unor modalități prin care clasificarea imaginilor poate fi utilizată în practica educațională.

5.6.2. Analiza sentimentelor folosind platforma Clarifai

5.6.2.1. Modelul ASI-C

Modelul ASI-C (Analiza Sentimentelor din Imagini – Clarifai) va clasifica imaginile în conformitate cu două tipuri de emoții, pozitive și negative. Pentru aceasta, am creat două concepte, POZITIV și NEGATIV. Scopul principal este de a studia performanțele modelului creat cu platforma Clarifai comparativ cu cele ale modelului ce efectuează clasificări similare, care utilizează AutoML Vision de la Google. Un alt obiectiv pe care l-am avut în vedere este de a identifica avantaje și dezavantaje ale utilizării de instrumente, arhitecturi, aplicații specifice pentru

lucrul cu Big Data, în format imagine, în manieră comparativă, încercând a face aprecieri asupra oportunității utilizării acestora pe scară largă pentru seturi mari de date provenite din educație.

Setul de date utilizat este *Imagini-generale*.

Scorul de acuratețe a clasificării (Model Accuracy Score) la nivelul celor două concepte, este de 0,84945774. Toate calculele au fost făcute pentru un prag de încredere (Threshold) de 0,5. Numărul de preziceri True Positives pentru clasa POZITIV este 11, cu o probabilitate mai mare sau egală decât 0,5. 6 imagini au fost etichetate POZITIV cu probabilități mai mici decât 0,5. Acestea constituie numărul de cazuri False Negatives. Numărul de cazuri False Positives este 0, ceea ce determină o valoare a Precision de 11/11, adică 1, mult mai mare decât precizia de clasificare a imaginilor cu eticheta NEGATIV, care este 0,727. În schimb valorile pentru Recall sunt destul de apropiate pentru ambele clase: 0,647 pentru POZITIV și 0,571 pentru NEGATIV (Fig. 23).

CONCEPT	K-SPLIT AVG	1 SPLIT						
	ACCURACY SCORE (ROC AUC)	TOTAL LABELED	TOTAL PREDICTED	TRUE POSITIVES	FALSE NEGATIVES	FALSE POSITIVES	RECALL RATE	PRECISION RATE
NEGATIV	0.811	14	11	8	6	3	0.571	0.727
POZITIV	0.888	17	11	11	6	0	0.647	1.000
TOTAL	AVG: 0.849	31	22	19	12	3	AVG: 0.609	AVG: 0.864

Figura 23. Rezultate obținute

În urma unei analize comparative privind performanțele de clasificare ale ASI-C față de modelul din Google Cloud, cu două etichete de clasă, ce folosește aceleași imagini, etichetate manual în același mod, se poate observa că cel din urmă este mai bun (Tabelul 6).

Tabel 6. Comparatie între rezultatele obținute de modelele ASI-G2 și ASI-C

	Model ASI-G2	Model ASI-C
Precizie medie	0,929	0,849
Precision la nivelul ambelor clase	0,875	0,609
Recall la nivelul ambelor clase	0,875	0,864
Precision pentru clasa POZITIV	0,888	1
Recall pentru clasa POZITIV	0,888	0,647
Precision pentru clasa NEGATIV	0,857	0,727
Recall pentru clasa NEGATIV	0,857	0,571

5.6.2.2. Modelul ASI-C-Edu

Am utilizat experiența unui model de tip *trainable*. Pentru a transfera învățarea către modelul personalizat ce urmează să fie antrenat pe cele două concepte noi, am folosit un clasificator bazat pe context (Context-Based Classifier).

Parametrul *output_info.output_config.concepts_mutually_exclusive* a fost setat cu valoarea False deoarece ambele concepte pot fi prezente într-o imagine, în proporții diferite, fiind vorba de concepte abstracte. Întrucât fiecare imagine din setul de date trebuie mapată la una din cele două categorii, valoarea parametrului *output_info.output_config.closed_environment* este True. Modelul

ar putea fi utilizat pentru a recunoaște context pozitiv sau negativ din imagini de pe camere de luat vederi, de pe o platformă educațională, etc.

Este utilizat setul de date *Imagini-Edu* (Capitolul 4).

Scorul de acuratețe (Model Accuracy Score), la un prag de încredere de 0.5 este de 0.711.

Pentru același prag de încredere, din cele 9 imagini folosite pentru evaluare, mapate conceptului NEGATIV, 5 sunt cazuri True Positive (au fost prezise corect ca fiind de tip negativ cu o probabilitate mai mare de 0.5) și 4 sunt False Negative. Pe baza acestora a fost calculată valoarea parametrului Recall, aceasta fiind 5/9, adică 55.6% (Tabelul 7).

Tabel 7. Rezultate obținute cu modelul ASI-C-Edu

CONCEPT	ACCURACY SCORE (ROC AUC)	TOTAL LABELED	TOTAL PREDICTED ▲	TRUE POSITIVES	FALSE NEGATIVES	FALSE POSITIVES	RECALL RATE	PRECISION RATE
negativ	0.758	9	6	5	4	1	0.556	0.833
pozitiv	0.664	9	6	4	5	2	0.444	0.667
TOTAL	AVG: 0.711	18	12	9	9	3	AVG: 0.500	AVG: 0.750

O analiză comparativă între performanțele de clasificare ale modelelor ASI-C și ASI-C-Edu relevă faptul că modelul ce folosește date din educație este mai slab comparativ cu celălalt (Tabelul 8).

Tabel 8. Comparatie între rezultatele obținute de modelele ASI-C-Edu și ASI-C

	Model ASI-C-Edu	Model ASI-C
Scorul de acuratețe	0,711	0,849
Precision pentru clasa POZITIV	0,667	1
Recall pentru clasa POZITIV	0,444	0,647
Precision pentru clasa NEGATIV	0,833	0,727
Recall pentru clasa NEGATIV	0,556	0,571

O posibilă explicație ar putea fi experiența modelului preantrenat să clasifice concepte concrete cum ar fi: peisaje, obiecte, culori deschise, nori. Imaginile din setul de date sunt destul de asemănătoare, de exemplu la lecțiile online, modelul ar trebui să identifice context pozitiv sau negativ având la dispoziție doar analiza fețelor. Același lucru este valabil și pentru setul de imagini preluate de la prezentarea proiectelor din timpul orei de informatică, unde modelul trebuie să clasifice bazându-se pe: identificarea unui zâmbet, a unei anumite poziții corporale, întrucât decorul este același.

Contribuțiile acestei cercetări constau în: realizarea unor proiecte de clasificare a imaginilor și formularea de recomandări privind alegerea celor mai potrivite soluții, în urma unor analize comparative.

5.6.3. Recomandări privind alegerea soluțiilor adecvate analizei sentimentelor elevilor referitoare la procesul educațional

Principală contribuție adusă de cercetările experimentale efectuate pentru clasificarea imaginilor este formularea unor concluzii și recomandări privind alegerea instrumentelor pentru

analiza sentimentelor, din punct de vedere al soluțiilor tehnice, dar și al cunoștințelor ce pot fi obținute din datele utilizate.

În urma analizei sentimentelor din imagini, folosind servicii specializate disponibile pe platformele Google și Clarifai, din punct de vedere al utilizării și performanțelor obținute de modele pot afirma următoarele:

- platforma Google oferă posibilitatea utilizării de modele care clasifică cu performanțe mai bune imaginile, dar prezintă dezavantajul unor costuri mai mari, proporționale cu mai mulți factori: cantitatea de date stocată, numărul de solicitări și altele, în timp ce platforma Clarifai oferă servicii gratuite pe o perioadă nedeterminată ce presupun acces la unele modele preantrenate, actualizate, fără costuri suplimentare. În plus, platforma Clarifai permite utilizarea cumulată a mai multor modele, maparea conceptelor proprii la un set predefinit prin intermediul grafului de cunoștințe, utilizarea unui sistem de adnotări cu ajutorul căruia se pot adăuga caracteristici suplimentare la conceptele deja create, pentru secțiuni din imagini și crearea și utilizarea propriilor modele;

- având în vedere performanțele mai scăzute ale unui model antrenat pe imagini obținute din diferite contexte educaționale, rămâne ca direcție viitoare de cercetare construirea și evaluarea unui model pentru seturi foarte mari de imagini, pentru a verifica dacă este posibilă o creștere a performanței acestora.

Din punct de vedere al cunoștințelor ce pot fi extrase din analiza datelor utilizate, am putut concluziona următoarele:

- folosind servicii specializate, care nu necesită cunoștințe avansate în domeniul învățării automate, am putut obține cu precizie destul de bună informații despre contextul pozitiv sau negativ al desfășurării unor activități. În vederea combaterii comportamentelor de umilire, denigratorii, a actelor de violență sau a infracțiunilor, poate fi realizat un sistem de alertă bazat pe modele care analizează imaginile preluate de camerele de supraveghere și identifică conceptele pozitiv/negativ.

- imaginile ce prezintă activități dinamice, sportive, desfășurate în natură sunt mapate cu precizie sporită conceptului pozitiv. Reacția pozitivă a elevilor la anumite activități și posibilitatea identificării la timp a acestora poate conduce la furnizarea de conținut atractiv, la metode diferite de expunere a conținuturilor.

- am observat că, modele cu aceeași arhitectură, dar antrenate cu alte seturi de imagini, produc rezultate foarte diferite. Astfel, un model antrenat cu imagini ce prezintă animale, natură, sporturi, obiecte obține rezultate mai bune decât atunci când este antrenat pe imagini din educație. Acest lucru se datorează faptului că am utilizat modele care folosesc o arhitectură dată și hiperparametri deja reglați, obișnuite să recunoască mai bine anumite lucruri din viața reală. În vederea obținerii de rezultate mai bune, devine oportună crearea de arhitecturi și modele proprii, ca parte a unor aplicații educaționale complexe, care să cuprindă și servicii specifice computer vision. În prezent, cercetările efectuate în domeniu, pentru sistemul preuniversitar românesc, sunt foarte puține, potențialul de explorare al acestui domeniu fiind important. Obținerea prototipului de elev face fezabilă instruirea personalizată a cărei eficiență nu poate fi contestată.

6. Recomandări privind direcțiile de acțiune în domeniul educației

Pe baza cercetărilor teoretice și practice efectuate și în urma studierii literaturii de specialitate privind direcțiile actuale de acțiune prevăzute în planurile pentru învățământ, naționale, locale și la nivelul școlii, am putut formula 6 direcții viitoare de acțiune pentru creșterea calității proceselor educaționale.

D1: *Elaborarea unei programe pentru un curs opțional dedicat elevilor de liceu (conținuturi, metode, procedee didactice, bibliografie, cadre și instrumente de lucru). Întocmirea documentelor și realizarea tuturor demersurilor necesare pentru ca aceasta să facă parte din oferta de cursuri opționale a Ministerului Educației.*

D2: *Elaborarea de programe de instruire pentru profesorii din învățământul preuniversitar pentru a dobândi competențe noi în materie de utilizare a noilor tehnologii informaționale și desfășurarea de activități de mentorat în parteneriat cu centrele universitare din țară.*

D3: *Desfășurarea de cursuri cu adulții în cadrul unor programe speciale, acreditate de Ministerul Educației și Ministerul Muncii, prin intermediul rețelei școlare din învățământul preuniversitar.*

D4: *Realizarea de platforme și aplicații de învățare, care să înglobeze tehnologii noi din domeniul explorării volumelor mari de date și care să cuprindă module pentru: învățare, evaluare, sisteme de recomandări, securitate, administrare și altele.*

D5: *Digitalizarea serviciilor din educație vizând componentele de administrație, management, instruire și evaluare și crearea unui cadru legislativ clar și coerent, a protocoalelor și metodologiilor specifice pentru preluarea și prelucrarea datelor provenite din sistem.*

D6: *Adaptarea statutului profesorului în școală și actualizarea fișei postului în conformitate cu necesitățile și problemele societății contemporane – crearea unor norme de cercetare și modificarea titularizării din forma actuală.*

Scopul acestor direcții de acțiune este de a intensifica colaborarea cu mediul universitar, dar și între consorțiile școlare și comunitatea locală, evidențiind încă o dată rolul educației ca principal pilon al unei societăți sănătoase.

7. Concluzii, contribuții personale și direcții viitoare de cercetare

7.1. Concluziile tezei

Cu scopul de a identifica cele mai potrivite modalități de utilizare a tehnologiilor de prelucrare a Big Data pentru creșterea calității serviciilor din educație, în special din sistemul preuniversitar, am efectuat o serie de cercetări experimentale constând în: procesarea textului cu ajutorul componentei MapReduce ca parte dintr-o arhitectură Hadoop Single Node Cluster 3.3.0, pe care am construit-o pe o mașină Ubuntu 20.04 LTS, utilizarea unui instrument de analitica datelor pentru a analiza comportamentul și tipologia utilizatorilor unui site educațional, investigație privind selecția tehnicilor de clasificare potrivite pentru a efectua predicții în legătură cu rezultatele obținute de elevi la Bacalaureat, realizarea unor proiecte de analiza a sentimentelor din date de tip text, folosind servicii specializate din Google Cloud pentru o abordare bazată pe învățare automată, respectiv instrumentul open-source Orange pentru o abordare bazată pe lexicon, ce clasifică sentimentele în funcție de emoțiile de bază descrise de psihologii Ekman și Plutchik, proiecte de analiza sentimentelor din imagini, folosind Google Cloud și platforma Clarifai, o analiză comparativă a serviciilor oferite de AWS și Google Cloud. În urma studiilor și experimentelor desfășurate, am putut trage concluzii valoroase: deși modelele antrenate folosind serviciul AutoML Vision din Google Cloud oferă rezultate mai bune în clasificarea imaginilor, platforma Clarifai este mai ofertantă din punct de vedere al flexibilității, transparenței și al costurilor de utilizare a

majorității serviciilor. Profilul de elev care promovează sau nu examenul de Bacalaureat rămâne neschimbat în perioada 2012-2019, care a fost analizată, ceea ce încurajează formularea unor măsuri procedurale care să îmbunătățească rezultatele, acestea având continuitate. Rezultatele clasificărilor efectuate sunt foarte bune, situate între 0.956 (SVM) și 1 (AdaBoost). Clasificarea opiniilor elevilor în legătură cu școala s-a făcut cu precizie destul de bună, anume 86,51%. Din păcate, peste 65% dintre elevi au opinii negative privind educația din țara noastră. Modelele antrenate pe concepte precum: natură, activități, animale, plante dau rezultate mai bune decât cele ce folosesc concepte abstracte precum teamă, tristețe, fericire, încredere sau negativ/pozitiv. Din acest motiv, se remarcă două direcții de cercetare: antrenarea unor modele, folosind seturi mari de imagini etichetate cu concepte abstracte pentru a studia performanța acestora și construirea propriilor modele în vederea integrării lor în aplicații educaționale, de exemplu pentru a furniza conținut media în funcție de comportamentul sau preferințele elevilor.

Pentru realizarea cercetărilor experimentale, am constituit o serie de seturi de date proprii ce au fost prezentate în capitolul 4. Pentru colectarea lor am utilizat: chestionare, rapoarte de analiză a interacțiunilor elevilor cu propriul site de instruire la informatică, imagini din diferite activități educative, aplicația informatică pentru Bacalaureat, depozit de date de pe platforma data.world. Pentru preprocesarea acestora am folosit: limbajul SQL, limbajul C++, instrumente pentru preprocesarea textului în vederea analizei sentimentelor prin abordarea bazată pe lexicon. O contribuție importantă este publicarea, în regim open acces, a unora dintre ele.

Având ca obiective identificarea unor noi direcții de cercetare, dar și formarea unei imagini corecte despre stadiul actual al cercetărilor în domeniul analizelor, tehnicilor și instrumentelor de prelucrare a Big Data din educație, am efectuat o analiză sistematică și critică a literaturii de specialitate. Căutarea sistemică a fost efectuată după metoda lui Kitchenham [7], în 4 baze de date: IEEE Xplore, Google Scholar, ScienceDirect și Web of Science. În urma aplicării criteriilor de includere și excludere pe care le-am stabilit, am selectat 38 de articole din 26 de domenii. Pentru analiza acestora am utilizat o abordare calitativă ce combină tehnicile meta-narativă și meta-analitică. Concluziile au permis formularea unor direcții de cercetare viitoare.

Pe baza cercetărilor teoretice și practice și în urma analizei sistematice a literaturii de specialitate, dar și consultării planurilor de acțiune din învățământ, am elaborat o serie de 6 direcții de acțiune care să contribuie la creșterea calității serviciilor din educație.

7.2. Contribuții personale

7.2.1. Contribuții teoretice

T1: Analiza sistematică și critică a literaturii de specialitate cu scopul de a aprecia stadiul actual al cercetărilor în domeniul Big Data din educație (Capitolul 2). Pentru a obține o imagine asupra stadiului actual al cercetărilor privind *folosirea instrumentelor și arhitecturilor specifice Big Data și aplicarea metodelor și tehnicilor de explorare a datelor din sistemul educațional românesc*, am efectuat o analiză a literaturii de specialitate. Procesul de căutare a fost unul sistematic, după metoda Kitchenham [7], în 4 baze de date: Google Scholar, Web of Science, IEEE Xplore și ScienceDirect, pentru perioada 2008-2018, rezultând un număr de 756 de articole publicate în jurnale sau prezentate la diverse conferințe și capitole de carte. În urma aplicării criteriilor de includere și excludere, am selectat 38 de studii. Pentru sinteza rezultatelor și formularea

concluziilor am utilizat o abordare calitativă care combină tehnica meta-analitică cu cea meta-narativă.

T2: Analiza comparativă a serviciilor cloud AWS și Google Cloud (Capitolul 3). Oportunitatea utilizării serviciilor cloud oferite de o firmă anume presupune cunoașterea avantajelor și a dezavantajelor față de oferta altor firme prezente pe piață. Comparația elaborată are în vedere aspecte precum: serviciile oferite, securitate, vizualizare, disponibilitate, rețea și altele. Ambele companii oferă servicii similare printre acestea numărându-se și cele pentru Big Data și învățare automată. Prin experimentarea ambelor variante, am ales să utilizez, în cercetările prezentate în lucrarea de doctorat serviciile Google Cloud.

T3: Colectarea, preprocesarea și constituirea a 6 seturi de date (Capitolul 4). În vederea realizării experimentelor practice, am colectat date de tipuri diferite: text, interacțiuni cu un site de învățare, imagini, numerice, folosind surse precum: grup țintă alcătuit din elevi de liceu, de la toate profilele și specializările din țară, Google Analytics, imagini personale, aplicația informatică folosită la examenul de Bacalaureat, platformă cu acces deschis la date. Acestea au fost preprocesate: validare, selectate folosind limbajul SQL, împărțirea în fișiere text folosind limbajul de programare C++, etichetare manuală, etc.

T4: Elaborarea unor direcții viitoare de acțiune (Capitolul 6). În urma studierii literaturii de specialitate și realizării unor studii de caz și experimente, am formulat 6 direcții de acțiune ce pot contribui la schimbarea unor paradigme, la actualizarea statutului profesorului în școală, la crearea de conținut și inovare, implicarea mediului de afaceri în realizarea unor activități de instruire și păstrarea relevanței în societate a angajaților prin educația de-a lungul întregii vieți, folosind infrastructura oferită de sistemul de învățământ preuniversitar.

7.2.2. Contribuții practice

P1: Crearea unei arhitecturi Hadoop Single Node Cluster 3.3.0 pe o mașină Ubuntu 20.04 LTS și realizarea unui studiu de caz ce utilizează componenta MapReduce (Capitolul 5). Am creat o arhitectură specifică pentru stocarea și procesarea Big Data, configurând totodată și o serie de componente (HDFS, YARN, MapReduce, JVM MapReduce, SSH) cu scopul de a realiza procesarea textului.

P2: Utilizarea Learning Analytics pentru analiza comportamentului și tipologiei utilizatorilor de site-uri educaționale (Capitolul 5). Am prezentat ca studiu de caz utilizarea analiticii învățării pentru îmbunătățirea calității serviciilor oferite de site-ul personal de pregătire la informatică. Analiza datelor din interacțiunile elevilor cu site-ul folosește instrumente specifice business intelligence și analiticii Web pentru a identifica comportamentul, tipurile și preferințele utilizatorilor, precum și distribuția geografică a acestora și tehnologiile și canalele folosite pentru conectarea la site. Remarcând interesul crescut pentru site din partea elevilor și profesorilor din toată țara, mi-am propus îmbunătățirea acestuia prin adăugarea de lecții, secțiuni de evaluare, mai multă interactivitate.

P3: Investigație privind selecția tehnicilor de clasificare pentru predicția rezultatelor la Bacalaureat (Capitolul 5). Folosind un set de date constituit din rezultatele obținute la examenul de Bacalaureat de către elevii Colegiului Economic "Dimitrie Cantemir" din Suceava, în perioada 2012-2019, am clasificat participanții în două categorii: PROMOVAT/RESPINS și am studiat corelațiile care există între rezultatele finale și anumite componente din date. Am utilizat algoritmii de clasificare: AdaBoost, SVM, k-NN, Neural Network și Naive Bayes și am comparat performanțele obținute de aceștia printr-o tehnică de validare încrucișată care folosește metricile de evaluare: AUC (Area Under the ROC Curve), CA (Classification Accuracy), F1, Precision și

Recall. Există o corelație slabă între rezultatele finale și cele obținute la competențele digitale și lingvistice. Profilul elevului promovat sau respins este același în perioada 2012-2019, ceea ce face oportună aplicarea pe termen lung, cu durată de cel puțin 10 ani, a acelorași procedee și metode de îmbunătățire a rezultatelor la bacalaureat.

P4: Implementarea proiectelor de analiză a sentimentelor folosind serviciul AutoML Language din Google Cloud (Capitolul 5). Am creat două proiecte de analiza sentimentelor folosind o abordare bazată pe învățare automată. Am utilizat două modele, ASG și Opinii, cu aceeași arhitectură, dar folosind seturi diferite de date pentru antrenare, validare și evaluare. Experiența acestora a fost folosită la etichetarea setului de date *191-opinii-despre-școală*. În scopul vizualizării rezultatelor, din consola Google Cloud, am conectat serviciul BigQuery pentru Big Data cu instrumentul Data Studio. Performanțele modelului Opinii sunt mult mai bune, comparativ cu ASG, ceea ce face mai credibilă concluzia că 65% dintre elevi au opinii negative în ceea ce privește școala. Concluziile acestui experiment sunt prezentate împreună cu o analiză comparativă a celor două modele privind performanțele de clasificare și informațiile pe care le-am obținut din date prin utilizarea lor.

P5: Analiza sentimentelor cu modelele ASO-Ekman și ASO-Plutchik (Capitolul 5). O abordare diferită de realizare a analizei sentimentelor utilizează două metode de clasificare a emoțiilor de bază elaborate de psihologii Ekman și Plutchik. Spre deosebire de studiile anterioare unde am utilizat algoritmi de învățare automată, în cercetarea prezentată în acest capitol am folosit o abordare bazată pe lexicon. Analizând rezultatele obținute cu toate modelele de analiză a textului, pot aprecia că abordarea bazată pe învățare automată oferă informații mai realiste în analiza sentimentelor elevilor în legătură cu școala.

P6: Proiectarea și implementarea proceselor de analiză a sentimentelor din imagini (Capitolul 5). Acest experiment înglobează, de fapt, mai multe cercetări privind clasificarea imaginilor pentru care am utilizat soluția AutoML Vision din Google Cloud și platforma Clarifai. Setul de date *Imagini-Generale* a fost folosit pentru antrenarea modelelor ASI-G4, ASI-G2 și ASI-C. Datele au trecut prin 3 procese de etichetare manuală: unul în care am utilizat 4 etichete de clasă, altul cu două etichete și un al treilea în care am creat și utilizat două concepte abstracte. Performanțele modelului ASI-G2 relevă faptul că este mai ușoară identificarea de context pozitiv sau negativ și că, prin adăugarea altor caracteristici abstracte în analiză, scade performanța de clasificare. Setul de date *Imagini-Edu* ce conține imagini personale din diferite activități școlare a fost folosit pentru antrenarea și evaluarea modelului ASI-C-Edu. Am constatat o scădere a performanțelor de clasificare. Analizele comparative efectuate au avut ca scop identificarea celei mai potrivite platforme pentru crearea și integrarea unor soluții de analiză a imaginilor în aplicațiile educative. Clarifai prezintă avantaje precum: gratuitatea majorității serviciilor, crearea propriilor modele, existența unui sistem de adnotări, maparea la concepte predefinite, folosirea cumulativă a experienței mai multor modele simultan.

7.3. Diseminarea rezultatelor

Contribuțiile personale și rezultatele obținute în cercetările desfășurate au fost prezentate la 5 conferințe internaționale și publicate în 9 jurnale și proceeding-uri ale conferințelor.

7.3.1. Conferințe internaționale

C1: BRAIN-BROAD RESEARCH IN ARTIFICIAL INTELLIGENCE AND NEUROSCIENCE, 2019, Sinaia, Brașov, România – Web of Science Core Collection: Emerging Sources Citation Index

C2: The 8th PHEADE International Conference (PHEADE 2019 – CCT), under the theme “Cultural Challenges of Technology”, October 17-19, Suceava, ROMÂNIA - **International Interdisciplinary Conference**

C3: 2020 International Conference on DEVELOPMENT AND APPLICATION SYSTEMS 15th Edition, May 21-23, Suceava, ROMANIA – **conferință IEEE, id: 49615, Proceedings in IEEE Xplore - WOS**

C4: ICERI2020 (13th annual International Conference of Education, Research and Innovation, 9th - 10th of November, 2020, Spain - virtually)

C5: INTED2021 (15th annual International Technology, Education and Development Conference, 8th-9th of March, 2021, Spain, Valencia - virtually)

7.3.2. Lista articolelor publicate în jurnale și proceeding-uri ale conferințelor

A1: Marcu, D. , Milici, D. L. , & Danubianu, M. (2020). *Software Engineering Ethics*. Postmodern Openings, 11(4), 248-261. <https://doi.org/10.18662/po/11.4/2334>. - **indexat Web of Science, JCI 0.73, Quartile Q2, Accession Number WOS:000600615200014**

A2: Marcu, Daniela; Danubianu, Mirela (2019). *Learning Analytics or Educational Data Mining? This is the Question....* BRAIN. Broad Research in Artificial Intelligence and Neuroscience, [S.l.], v. 10, p. 1-14, oct. 2019. ISSN 2067-3957. <https://www.edusoft.ro/brain/index.php/brain/article/view/953> - **indexat Web of Science, JCI 0.16, Quartile Q4, Accession Number WOS:000491308800001**

A3: D. Marcu and M. Danubianu, (2020). *Sentiment Analysis from Students' Feedback : A Romanian High School Case Study*. 2020 International Conference on Development and Application Systems (DAS), Suceava, Romania, 2020, pp. 204-209, doi: 10.1109/DAS49615.2020.9108927. - **indexat Web of Science, Accession Number WOS:000589776100037**

A4: Marcu, Daniela; Danubianu, Mirela; Pentiuc, Gheorghe (2019). *Big Data Technology - valuable tool for healthcare research*. International Journal of Computer Science and Network Security, April 30, 2019. ISSN : 1738-7906. Vol.19, No.4. http://ijcsns.org/07_book/html/201904/201904016.html – **indexat Web of Science, JCI 0.09, Quartile Q4, Accession Number WOS:000472817600016**

A5: D. Marcu, M. Danubianu, A. Barîlă, C. Simionescu (2021). *Algorithms for Classifying the Results at the Baccalaureate Exam - Comparative Analysis of Performances*. International Journal of Computer Science and Network Security ISSN : 1738-7906. Volume Number: Vol.21, No.05. http://ijcsns.org/07_book/html/202108/202108005.html, DOI: 10.22937/IJCSNS.2021.21.8.5 1 – **indexat Web of Science, JCI 0.09, Quartile Q4, Accession Number WOS:000697025200005**

A6: D. Marcu, M. Danubianu (2021). *Sentiment Analysis From Images - Comparative Study of SAI-G and SAI-C Models' Performances Using AutoML Vision Service from Google Cloud and Clarifai Platform*. International Journal of Computer Science and Network Security, Vol. 21, No. 9, pp. 179-184. ISSN 1738-7906. http://ijcsns.org/07_book/html/202109/202109025.html, DOI: 10.22937/IJCSNS.2021.21.9.25 – **indexat Web of Science, JCI 0.09, Quartile Q4, Accession Number WOS:000708805900025**

A7: Corina Simionescu, Mirela Danubianu, **Daniela Marcu**, Corneliu-Octavian Turcu (2021). *Online Learning after One Year of Digital Schooling in Romania: A Survey*. IJCSNS International Journal of Computer Science and Network Security, VOL.21 No.12, December 2021. DOI: 10.22937/IJCSNS.2022.22.1.5 – **indexat Web of Science, JCI 0.09, Quartile Q4, Accession Number WOS:000755171400005**

A8: C. Simionescu, M. Danubianu, **D. Marcu** (2020). *Analysis of online education Romanian schools due to Covid-19 pandemics and areas of improvement*, ICERI2020 Proceedings, pp. 3523-3529. DOI: doi: 10.21125/iceri.2020.0787.

A9: **D. Marcu**, M. Danubianu, C. Simionescu (2021). *Comparative analysis of predictive models on online education in context of Covid-19 – A case study*, INTED2021 Proceedings, pp. 4403-4412. DOI: 10.21125/inted.2021.0899.

7.3.3. Implicarea în proiecte de cercetare

Proiect, perioada 1 septembrie 2020 – 31 august 2021: *Exceleța academică și valori antreprenoriale - sistem de burse pentru asigurarea oportunităților de formare și dezvoltare a competențelor antreprenoriale ale doctoranzilor și postdoctoranzilor (ANTREPRENORDOC)*, cofinanțat din Fondul Social European prin Programul Operațional Capital Uman, 2014-2020, Contract nr. 36355/23.05.2019 POCU/380/6/13 - Cod SMIS: 123847

7.4. Direcții viitoare de cercetare

În urma cercetărilor teoretice și a studiilor experimentale efectuate, a concluziilor pozitive sau negative pe care le-am tras, am identificat o serie de direcții viitoare de cercetare și de acțiune:

1. Crearea de platforme moderne de învățare dedicate învățământului preuniversitar care să înglobeze tehnologii noi și servicii specifice din domenii precum: prelucrarea limbajului natural, computer vision, analitica datelor, Big Data.

2. Achiziționarea de sisteme pentru stocarea și procesarea Big Data, la nivel experimental, pentru grupuri de școli pilot, cu scopul de a pregăti cercetători în domenii specifice care să fie capabili să utilizeze aceste arhitecturi pentru a moderniza departamente din cadrul sistemului educațional preuniversitar.

3. Efectuarea analizei sentimentelor pentru texte în limba română: crearea de instrumente, algoritmi specifici, dicționare.

4. Crearea de cursuri opționale de: Big Data, metodica cercetării, învățare automată, computer vision pentru elevii de liceu și introducerea acestora în programul național de cursuri opționale acreditate de Ministerul Educației împreună cu crearea unor programe de specializare pentru profesorii de informatică și TIC care să poată preda aceste cursuri, în colaborare cu universitățile din țară.

5. Realizarea unei analize a literaturii de specialitate care să cuprindă cercetări și studii publicate în ultimii 5 ani ce vizează Big Data din educația românească și studierea rezultatelor, în manieră comparativă cu cele obținute prin analiza literaturii din perioada 2008-2018.

Acknowledgment

„Această lucrare a fost realizată în cadrul proiectului *Exceleța academică și valori antreprenoriale - sistem de burse pentru asigurarea oportunităților de formare și dezvoltare a competențelor antreprenoriale ale doctoranzilor și postdoctoranzilor (ANTREPRENORDOC)*, cofinanțat din Fondul Social European prin Programul Operațional Capital Uman, 2014-2020, Contract nr. 36355/23.05.2019 POCU/380/6/13 - Cod SMIS: 123847.”

„This work is supported by the project *ANTREPRENORDOC*, in the framework of Human Resources Development Operational Programme 2014-2020, financed from the European Social Fund under the contract number 36355/23.05.2019 HRD OP /380/6/13 – SMIS Code: 123847.”

Bibliografie

- [1] Baumeister, R. F., & Leary, M. R. (1997). *Writing narrative literature reviews*. Review of general psychology, 1(3), pp. 311-320.
- [2] Hannah Snyder (2019). *Literature review as a research methodology: An overview and guidelines*. Journal of Business Research, Volume 104, pp. 333-339, ISSN 0148-2963, <https://doi.org/10.1016/j.jbusres.2019.07.039>.
- [3] Torraco, R.J. (2005). *Writing Integrative Literature Reviews: Guidelines and Examples*. Human Resource Development Review, 4, pp. 356-367. DOI: <http://dx.doi.org/10.1177/1534484305278283>
- [4] Torres-Carrion, Pablo & González González, Carina & Aciar, Silvana & Rodriguez, Germana. (2018). *Methodology for Systematic Literature Review applied to Engineering and Education*. DOI: 10.1109/EDUCON.2018.8363388.
- [5] Greenhalgh, Trisha & Robert, Glenn & Macfarlane, Fraser & Bate, Paul & Kyriakidou, Olivia. (2004). *Diffusion of Innovations in Service Organizations: Systematic Review and Recommendations*. The Milbank quarterly, 82, pp. 581-629. DOI: 10.1111/j.0887-378X.2004.00325.x.
- [6] Wong, Geoff & Greenhalgh, Trisha & Westhorp, Gill & Buckingham, Jeanette & Pawson, Ray. (2013). *RAMESES publication standards: Meta-narrative reviews*. BMC medicine. 11. 20. 10.1186/1741-7015-11-20.
- [7] Kitchenham, Barbara. (2004). *Procedures for Performing Systematic Reviews*. Keele, UK, Keele Univ., 33.
- [8] YILDIRIM, P., & Birant, D. (2018). Bulut bilişimde veri madenciliği tekniklerinin uygulanması: Bir literatür taraması. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 24(2), 336-343.
- [9] Eroglu, Y., & Seçkiner, S. U. (2016). Trend Topic Analysis for Wind Energy Researches: A Data Mining Approach Using Text Mining. *Journal of Technology Innovations in Renewable Energy*, 5(2), 44-58.
- [10] Mohassel, R. R., Fung, A. S., Mohammadi, F., & Raahemifar, K. (2014, May). A survey on advanced metering infrastructure and its application in smart grids. In *2014 IEEE 27th Canadian conference on electrical and computer engineering (CCECE)* (pp. 1-8). IEEE.
- [11] Dimova, D. (2018). Application of statistical methods for studying time series on trips of Bulgarian residents abroad. *International Multidisciplinary Scientific GeoConference: SGEM*, 18(2.1), pp. 141-148.
- [12] Yukino Baba, Kei Kinoshita, Hisashi Kashima (2016). Participation recommendation system for crowdsourcing contests. *Expert Systems with Applications*, Volume 58, 2016, pp. 174-183, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2016.04.010>.
- [13] Jaesung Lee, Dae-Won Kim (2016). An effective initialization method for genetic algorithm-based robot path planning using a directed acyclic graph. *Information Sciences*, Volume 332, 2016, pp. 1-18, ISSN 0020-0255, <https://doi.org/10.1016/j.ins.2015.11.004>.
- [14] Enmei Tu, Longbing Cao, Jie Yang, Nicola Kasabov (2014). A novel graph-based k-means for nonlinear manifold clustering and representative selection. *Neurocomputing*, Volume 143, 2014, pp. 109-122, ISSN 0925-2312, <https://doi.org/10.1016/j.neucom.2014.05.067>.
- [15] Ianculescu, M; Alexandru, A and Gheorghe-Moisii, M. (2017). *Harnessing the Potential of Big Data in Romanian Healthcare*. 2017 5th International Symposium On Electrical and Electronics Engineering (ISEEE). ISSN2378-3907
- [16] Geman, O; Chiuchisan, I; (...); Milici, LD. (2018). *Deep Learning Tools for Human Microbiome Big Data*. 2018 | Soft Computing Applications, SOFA 2016, VOL 1 633 , pp.265-275. DOI10.1007/978-3-319-62521-8_21
- [17] A. Garcia-Saavedra, G. Iosifidis, X. Costa-Perez and D. J. Leith (2018). *Joint Optimization of Edge Computing Architectures and Radio Access Networks*, in IEEE Journal on Selected Areas in Communications, vol. 36, no. 11, pp. 2433-2443, Nov. 2018, doi: 10.1109/JSAC.2018.2874142.

- [18] O. Cardin, F. Ounnar, A. Thomas and D. Trentesaux (2017). *Future Industrial Systems: Best Practices of the Intelligent Manufacturing and Services Systems (IMS2) French Research Group*, in IEEE Transactions on Industrial Informatics, vol. 13, no. 2, pp. 704-713, April 2017, doi: 10.1109/TII.2016.2605624.
- [19] Tian, Y., Shi, Y., & Liu, X. (2012). Recent advances on support vector machines research. *Technological and economic development of Economy*, 18(1), 5-33.
- [20] Yildirim, P., & Birant, D. (2018). Application of data mining techniques in cloud computing: A literature review. *Pamukkale University Journal of Engineering Sciences-Pamukkale Universitesi Muhendislik Bilimleri Dergisi*, 24(2).
- [21] Lima, N., Viegas, C., Alves, G., & Garcia-Peñalvo, F. J. (2016). VISIR's Usage as an Educational Resource: a Review of the Empirical Research. *Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality*, pp. 893-901.
- [22] García-Peñalvo, F. J., Alves, G. R., Lima, N., & Viegas, C. (2016). *VISIR's Usage as an Educational Resource: a Review of the Empirical Research*.
- [23] Cong, J., & Liu, H. (2014). Approaching human language with complex networks. *Physics of life reviews*, 11(4), pp. 598-618.
- [24] Pareja-Lora, A., Rodríguez-Arancón, P., & Calle-Martínez, C. (2016). Applying information and communication technologies to language teaching and research: an overview. *New perspectives on teaching and working with languages in the digital era*, pp. 1-22.
- [25] Salgado, J. F. (2017). Using ability tests in selection. *The Wiley Blackwell Handbook of the Psychology of Recruitment, Selection and Employee Retention*, pp. 113-150.
- [26] Bogoslov, I. A. (2018). Future research directions on web-based educational systems. In *International Economic Conference of Sibiu*, pp. 9-20. Springer, Cham.
- [27] Busoniu, L., Babuska, R., & De Schutter, B. (2008). A comprehensive survey of multiagent reinforcement learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 38(2), pp. 156-172.
- [28] Dascălu, E. D., Duguleană, L., Duguleană, C., Busuiocanu, S., Dragomir, I. R., Lazăr, M., & Davidescu, P. A. (2016). *Romanian statistical review*, Nr. 3/2016. Online: www.revistadestatistica.ro.
- [29] Museros, L., Pujol, O., & Agell, N. (Eds.). (2014). *Artificial Intelligence Research and Development: Recent Advances and Applications*, IOS Press, ISBN-10 161499451X.
- [30] Papamitsiou, Anastasios A. (2017). Economides, exhibiting achievement behavior during computer-based testing: What temporal trace data and personality traits tell us? *Computers in Human Behavior*, Volume 75, pp. 423-438, ISSN 0747-5632, <https://doi.org/10.1016/j.chb.2017.05.036>.
- [31] Jorge Maldonado-Mahauad, Mar Pérez-Sanagustín, René F. Kizilcec, Nicolás Morales, Jorge Munoz-Gama (2018). Mining theory-based patterns from Big data: Identifying self-regulated learning strategies in Massive Open Online Courses. *Computers in Human Behavior*, Volume 80, 2018, pp. 179-196, ISSN 0747-5632, <https://doi.org/10.1016/j.chb.2017.11.011>.
- [32] Jorge Miguel, Santi Caballé, Fatos Xhafa, Chapter 4 - Trustworthiness modeling and methodology for secure peer-to-peer e-Assessment, Editor(s): Jorge Miguel, Santi Caballé, Fatos Xhafa, In *Intelligent Data-Centric Systems, Intelligent Data Analysis for e-Learning*, Academic Press, 2017, pp. 49-81, ISBN 9780128045350, <https://doi.org/10.1016/B978-0-12-804535-0.00004-6>.
- [33] Vitomir Kovanović, Dragan Gašević, Srećko Joksimović, Marek Hatala, Olusola Adesope, Analytics of communities of inquiry: Effects of learning technology use on cognitive presence in asynchronous online discussions, *The Internet and Higher Education*, Volume 27, 2015, Pages 74-89, ISSN 1096-7516, <https://doi.org/10.1016/j.iheduc.2015.06.002>.
- [34] Susanne Narciss, Sergey Sosnovsky, Lenka Schnaubert, Eric Andrès, Anja Eichelmann, George Goguadze, Erica Melis (2014). *Exploring feedback and student characteristics relevant for personalizing feedback strategies*,

Computers & Education, Volume 71, 2014, Pages 56-76, ISSN 0360-1315, <https://doi.org/10.1016/j.compedu.2013.09.011>.

[35] Cristóbal Romero, Sebastián Ventura, Amelia Zafra, Paul de Bra (2009). Applying Web usage mining for personalizing hyperlinks in Web-based adaptive educational systems, *Computers & Education*, Volume 53, Issue 3, 2009, Pages 828-840, ISSN 0360-1315, <https://doi.org/10.1016/j.compedu.2009.05.003>.

[36] Sarah K. Howard, Jun Ma, Jie Yang (2016). Student rules: Exploring patterns of students' computer-efficacy and engagement with digital technologies in learning, *Computers & Education*, Volume 101, 2016, Pages 29-42, ISSN 0360-1315, <https://doi.org/10.1016/j.compedu.2016.05.008>.

[37] Maria Theobald, Henrik Bellhäuser, Margarete Imhof (2018). Identifying individual differences using log-file analysis: Distributed learning as mediator between conscientiousness and exam grades. *Learning and Individual Differences*, Volume 65, 2018, Pages 112-122, ISSN 1041-6080, <https://doi.org/10.1016/j.lindif.2018.05.019>.

[38] Roger Nkambou, Philippe Fournier-Viger, Engelbert Mephu Nguifo (2011). Learning task models in ill-defined domain using a hybrid knowledge discovery framework, *Knowledge-Based Systems*, Volume 24, Issue 1, 2011, Pages 176-185, ISSN 0950-7051, <https://doi.org/10.1016/j.knosys.2010.08.002>.

[39] Farshid Marbouti, Heidi A. Diefes-Dux, Krishna Madhavan (2016). Models for early prediction of at-risk students in a course using standards-based grading, *Computers & Education*, Volume 103, 2016, Pages 1-15, ISSN 0360-1315, <https://doi.org/10.1016/j.compedu.2016.09.005>.

[40] Benjamin Paassen, Bassam Mokbel, Barbara Hammer (2016). Adaptive structure metrics for automated feedback provision in intelligent tutoring systems, *Neurocomputing*, Volume 192, 2016, Pages 3-13, ISSN 0925-2312, <https://doi.org/10.1016/j.neucom.2015.12.108>.

[41] Samuel Greiff, Christoph Niepel, Ronny Scherer, Romain Martin (2016). Understanding students' performance in a computer-based assessment of complex problem solving: An analysis of behavioral data from computer-generated log files, *Computers in Human Behavior*, Volume 61, 2016, Pages 36-46, ISSN 0747-5632, <https://doi.org/10.1016/j.chb.2016.02.095>.

[42] Iria Estévez-Ayres, Jesús Arias Fisteus, Carlos Delgado-Kloos, Lostrego: A distributed stream-based infrastructure for the real-time gathering and analysis of heterogeneous educational data, *Journal of Network and Computer Applications*, Volume 100, 2017, Pages 56-68, ISSN 1084-8045, <https://doi.org/10.1016/j.jnca.2017.10.014>.

[43] Jelena Jovanović, Dragan Gašević, Shane Dawson, Abelardo Pardo, Negin Mirriahi (2017). Learning analytics to unveil learning strategies in a flipped classroom, *The Internet and Higher Education*, Volume 33, 2017, Pages 74-85, ISSN 1096-7516, <https://doi.org/10.1016/j.iheduc.2017.02.001>.

[44] M. Munk, M. Drlík (2016). Chapter 10 - Methodology of Predictive Modeling of Students' Behavior in Virtual Learning Environment, Editor(s): Santi Caballé, Robert Clarisó, In *Intelligent Data-Centric Systems, Formative Assessment, Learning Data Analytics and Gamification*, Academic Press, 2016, Pages 187-216, ISBN 9780128036372, <https://doi.org/10.1016/B978-0-12-803637-2.00010-5>.

[45] René F. Kizilcec, Mar Pérez-Sanagustín, Jorge J. Maldonado (2017). Self-regulated learning strategies predict learner behavior and goal attainment in Massive Open Online Courses, *Computers & Education*, Volume 104, 2017, Pages 18-33, ISSN 0360-1315, <https://doi.org/10.1016/j.compedu.2016.10.001>.

[46] Curtis R. Henrie, Lisa R. Halverson, Charles R. Graham (2015). Measuring student engagement in technology-mediated learning: A review, *Computers & Education*, Volume 90, 2015, Pages 36-53, ISSN 0360-1315, <https://doi.org/10.1016/j.compedu.2015.09.005>.

[47] Tao Xie, Qinghua Zheng, Weizhan Zhang (2015). Mining temporal characteristics of behaviors from interval events in e-learning, *Information Sciences*, Volume 447, 2018, Pages 169-185, ISSN 0020-0255, <https://doi.org/10.1016/j.ins.2018.03.018>.

[48] Cristóbal Romero, Manuel-Ignacio López, Jose-María Luna, Sebastián Ventura (2013). Predicting students' final performance from participation in on-line discussion forums, *Computers & Education*, Volume 68, 2013, Pages 458-472, ISSN 0360-1315, <https://doi.org/10.1016/j.compedu.2013.06.009>.

- [49] Fátima Rodrigues, Paulo Oliveira (2014). A system for formative assessment and monitoring of students' progress, *Computers & Education*, Volume 76, 2014, Pages 30-41, ISSN 0360-1315, <https://doi.org/10.1016/j.compedu.2014.03.001>.
- [50] Dan Davis, Guanliang Chen, Claudia Hauff, Geert-Jan Houben (2018). Activating learning at scale: A review of innovations in online learning strategies, *Computers & Education*, Volume 125, 2018, Pages 327-344, ISSN 0360-1315, <https://doi.org/10.1016/j.compedu.2018.05.019>.
- [51] Samuel Greiff, Sascha Wüstenberg, Francesco Avvisati (2015). Computer-generated log-file analyses as a window into students' minds? A showcase study based on the PISA 2012 assessment of problem solving, *Computers & Education*, Volume 91, 2015, Pages 92-105, ISSN 0360-1315, <https://doi.org/10.1016/j.compedu.2015.10.018>.
- [52] Jorge Miguel, Santi Caballé, Fatos Xhafa (2017). Chapter 3 - Trustworthiness for secure collaborative learning, Editor(s): Jorge Miguel, Santi Caballé, Fatos Xhafa, In *Intelligent Data-Centric Systems, Intelligent Data Analysis for e-Learning*, Academic Press, 2017, Pages 25-48, ISBN 9780128045350, <https://doi.org/10.1016/B978-0-12-804535-0.00003-4>.
- [53] Marija Blagojević, Živadin Micić (2013). A web-based intelligent report e-learning system using data mining techniques, *Computers & Electrical Engineering*, Volume 39, Issue 2, 2013, Pages 465-474, ISSN 0045-7906, <https://doi.org/10.1016/j.compeleceng.2012.09.011>.
- [54] Ilker Yengin, Adem Karahoca, Dilek Karahoca (2011). E-learning success model for instructors' satisfactions in perspective of interaction and usability outcomes, *Procedia Computer Science*, Volume 3, 2011, Pages 1396-1403, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2011.01.021>.
- [55] Barış Yüce, Adem Karahoca, Dilek Karahoca (2011). The use of electronic curriculums in occupational education to evaluate and improve the cognitive capacity of candidate software engineers, *Procedia Computer Science*, Volume 3, 2011, Pages 1418-1424, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2011.01.024>.
- [56] M. Mayilvaganan, D. Kalpanadevi (2015). Cognitive Skill Analysis for Students through Problem Solving Based on Data Mining Techniques, *Procedia Computer Science*, Volume 47, 2015, Pages 62-75, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2015.03.184>.
- [57] Sunita B. Aher, L.M.R.J. Lobo (2013). Combination of machine learning algorithms for recommendation of courses in E-Learning System based on historical data, *Knowledge-Based Systems*, Volume 51, 2013, Pages 1-14, ISSN 0950-7051, <https://doi.org/10.1016/j.knosys.2013.04.015>.
- [58] Daniel Alencar, Darlinton Carvalho, Eduardus Koenders, Fernando Mourão, Leonardo Rocha (2017). Devising a computational model based on data mining techniques to predict concrete compressive strength, *Procedia Computer Science*, Volume 108, 2017, Pages 455-464, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2017.05.018>.
- [59] Azer Nouira, Lilia Cheniti-Belcadhi, Rafik Braham (2016). An Enhanced xAPI Data Model Supporting Assessment Analytics, *Procedia Computer Science*, Volume 126, 2018, Pages 566-575, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2018.07.291>.
- [60] Ivana Ognjanovic, Dragan Gasevic, Shane Dawson (2016). Using institutional data to predict student course selections in higher education, *The Internet and Higher Education*, Volume 29, 2016, Pages 49-62, ISSN 1096-7516, <https://doi.org/10.1016/j.iheduc.2015.12.002>.
- [61] Xavi Canaleta, David Vernet, Lluís Vicent, José Antonio Montero (2014). Master in Teacher Training: A real implementation of Active Learning, *Computers in Human Behavior*, Volume 31, 2014, Pages 651-658, ISSN 0747-5632, <https://doi.org/10.1016/j.chb.2013.09.020>.
- [62] Danial Hooshyar, Rodina Binti Ahmad, Moslem Yousefi, Moein Fathi, Shi-Jinn Horng, Heuseok Lim (2016). Applying an online game-based formative assessment in a flowchart-based intelligent tutoring system for improving problem-solving skills, *Computers & Education*, Volume 94, 2016, Pages 18-36, ISSN 0360-1315, <https://doi.org/10.1016/j.compedu.2015.10.013>.

- [63] Ji Won You (2016). Identifying significant indicators using LMS data to predict course achievement in online learning, *The Internet and Higher Education*, Volume 29, 2016, Pages 23-30, ISSN 1096-7516, <https://doi.org/10.1016/j.iheduc.2015.11.003>.
- [64] Olga Viberg, Mathias Hatakka, Olof Bälter, Anna Mavroudi (2018). The current landscape of learning analytics in higher education, *Computers in Human Behavior*, Volume 89, 2018, Pages 98-110, ISSN 0747-5632, <https://doi.org/10.1016/j.chb.2018.07.027>.
- [65] Ya-Han Hu, Chia-Lun Lo, Sheng-Pao Shih (2014). Developing early warning systems to predict students' online learning performance, *Computers in Human Behavior*, Volume 36, 2014, Pages 469-478, ISSN 0747-5632, <https://doi.org/10.1016/j.chb.2014.04.002>.
- [66] Zhicheng Liao, Yangyong Zhu (2014). When a Classifier Meets More Data, *Procedia Computer Science*, Volume 30, 2014, Pages 50-59, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2014.05.380>.
- [67] Kathy Robinson (2013). The interrelationship of emotion and cognition when students undertake collaborative group work online: An interdisciplinary approach, *Computers & Education*, Volume 62, 2013, Pages 298-307, ISSN 0360-1315, <https://doi.org/10.1016/j.compedu.2012.11.003>.
- [68] Siti Khadijah Mohamad, Zaidatun Tasir (2013). Educational Data Mining: A Review, *Procedia - Social and Behavioral Sciences*, Volume 97, 2013, Pages 320-324, ISSN 1877-0428, <https://doi.org/10.1016/j.sbspro.2013.10.240>.
- [69] Janice D. Gobert, Yoon Jeon Kim, Michael A. Sao Pedro, Michael Kennedy, Cameron G. Betts (2015). Using educational data mining to assess students' skills at designing and conducting experiments within a complex systems microworld, *Thinking Skills and Creativity*, Volume 18, 2015, Pages 81-90, ISSN 1871-1871, <https://doi.org/10.1016/j.tsc.2015.04.008>.
- [70] Evandro B. Costa, Baldoino Fonseca, Marcelo Almeida Santana, Fabrícia Ferreira de Araújo, Joilson Rego (2017). Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses, *Computers in Human Behavior*, Volume 73, 2017, Pages 247-256, ISSN 0747-5632, <https://doi.org/10.1016/j.chb.2017.01.047>.
- [71] Wanli Xing, Rui Guo, Eva Petakovic, Sean Goggins (2015). Participation-based student final performance prediction model through interpretable Genetic Programming: Integrating learning analytics, educational data mining and theory, *Computers in Human Behavior*, Volume 47, 2015, Pages 168-181, ISSN 0747-5632, <https://doi.org/10.1016/j.chb.2014.09.034>.
- [72] Shinichi Oeda, Genki Hashimoto (2017). Log-Data Clustering Analysis for Dropout Prediction in Beginner Programming Classes, *Procedia Computer Science*, Volume 112, 2017, Pages 614-621, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2017.08.088>.
- [73] Nguyen Thai-Nghe, Lucas Drumond, Artus Krohn-Grimberghe, Lars Schmidt-Thieme (2010). Recommender system for predicting student performance, *Procedia Computer Science*, Volume 1, Issue 2, 2010, Pages 2811-2819, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2010.08.006>.
- [74] Rafi Nachmias (2011). Web mining and higher education: Introduction to the special issue, *The Internet and Higher Education*, Volume 14, Issue 2, 2011, Pages 65-66, ISSN 1096-7516, <https://doi.org/10.1016/j.iheduc.2011.01.001>.
- [75] Amirah Mohamed Shahiri, Wahidah Husain, Nur'aini Abdul Rashid (2015). A Review on Predicting Student's Performance Using Data Mining Techniques, *Procedia Computer Science*, Volume 72, 2015, Pages 414-422, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2015.12.157>.
- [76] M.T. Irfan, V.N. Gudivada (2016). Chapter 9 - Cognitive Computing Applications in Education and Learning, Editor(s): Venkat N. Gudivada, Vijay V. Raghavan, Venu Govindaraju, C.R. Rao, *Handbook of Statistics*, Elsevier, Volume 35, 2016, Pages 283-300, ISSN 0169-7161, ISBN 9780444637444, <https://doi.org/10.1016/bs.host.2016.07.008>.

- [77] V.L. Miguéis, Ana Freitas, Paulo J.V. Garcia, André Silva (2018). Early segmentation of students according to their academic performance: A predictive modelling approach, *Decision Support Systems*, Volume 115, 2018, Pages 36-51, ISSN 0167-9236, <https://doi.org/10.1016/j.dss.2018.09.001>.
- [78] Konomu Dobashi (2017). Automatic data integration from Moodle course logs to pivot tables for time series cross section analysis, *Procedia Computer Science*, Volume 112, 2017, Pages 1835-1844, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2017.08.222>.
- [79] Renza Campagni, Donatella Merlini, Renzo Sprugnoli, Maria Cecilia Verri (2015). Data mining models for student careers, *Expert Systems with Applications*, Volume 42, Issue 13, 2015, Pages 5508-5521, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2015.02.052>.
- [80] Satrio Adi Priyambada, E.R. Mahendrawathi, Bernardo Nugroho Yahya (2017). Curriculum Assessment of Higher Educational Institution Using Aggregate Profile Clustering, *Procedia Computer Science*, Volume 124, 2017, Pages 264-273, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2017.12.155>.
- [81] S. Kotsiantis, K. Patriarcheas, M. Xenos (2010). A combinational incremental ensemble of classifiers as a technique for predicting students' performance in distance education, *Knowledge-Based Systems*, Volume 23, Issue 6, 2010, Pages 529-535, ISSN 0950-7051, <https://doi.org/10.1016/j.knosys.2010.03.010>.
- [82] Sarah Mohamed Hassan, Muna S. Al-Razgan (2016). Pre-University Exams Effect on Students GPA: A Case Study in IT Department, *Procedia Computer Science*, Volume 82, 2016, Pages 127-131, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2016.04.018>.
- [83] Mudasir Ashraf, Majid Zaman, Muheet Ahmed (2018). Using Ensemble StackingC Method and Base Classifiers to Ameliorate Prediction Accuracy of Pedagogical Data, *Procedia Computer Science*, Volume 132, 2018, Pages 1021-1040, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2018.05.018>.
- [84] Thomas F. Stahovich, Hanlung Lin (2016). Enabling data mining of handwritten coursework, *Computers & Graphics*, Volume 57, 2016, Pages 31-45, ISSN 0097-8493, <https://doi.org/10.1016/j.cag.2016.01.002>.
- [85] Ghada Badr, Afnan Algobail, Hanadi Almutairi, Manal Almutery (2016). Predicting Students' Performance in University Courses: A Case Study and Tool in KSU Mathematics Department, *Procedia Computer Science*, Volume 82, 2016, Pages 80-89, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2016.04.012>.
- [86] Leyla Zhuhadar, Scarlett Marklin, Evelyn Thrasher, Miltiadis D. Lytras (2016). Is there a gender difference in interacting with intelligent tutoring system? Can Bayesian Knowledge Tracing and Learning Curve Analysis Models answer this question?, *Computers in Human Behavior*, Volume 61, 2016, Pages 198-204, ISSN 0747-5632, <https://doi.org/10.1016/j.chb.2016.02.073>.
- [87] Gökhan Akçapınar (2015). How automated feedback through text mining changes plagiaristic behavior in online assignments, *Computers & Education*, Volume 87, 2015, Pages 123-130, ISSN 0360-1315, <https://doi.org/10.1016/j.compedu.2015.04.007>.
- [88] Sandhya Maitra, Sushila Madan, Rekha Kandwal, Perna Mahajan (2018). Mining authentic student feedback for faculty using Naïve Bayes classifier, *Procedia Computer Science*, Volume 132, 2018, Pages 1171-1183, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2018.05.032>.
- [89] Guillaume Durand, Nabil Belacel, François LaPlante (2013). Graph theory based model for learning path recommendation, *Information Sciences*, Volume 251, 2013, Pages 10-21, ISSN 0020-0255, <https://doi.org/10.1016/j.ins.2013.04.017>.
- [90] Amelia Zafra, Cristóbal Romero, Sebastián Ventura (2011). Multiple instance learning for classifying students in learning management systems, *Expert Systems with Applications*, Volume 38, Issue 12, 2011, Pages 15020-15031, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2011.05.044>.
- [91] Juan A. Lara, David Lizcano, María A. Martínez, Juan Pazos, Teresa Riera (2014). A system for knowledge discovery in e-learning environments within the European Higher Education Area – Application to student data from Open University of Madrid, UDIMA, *Computers & Education*, Volume 72, 2014, Pages 23-36, ISSN 0360-1315, <https://doi.org/10.1016/j.compedu.2013.10.009>.

- [92] Arnon Hershkovitz, Rafi Nachmias (2011). Online persistence in higher education web-supported courses, *The Internet and Higher Education*, Volume 14, Issue 2, 2011, Pages 98-106, ISSN 1096-7516, <https://doi.org/10.1016/j.iheduc.2010.08.001>.
- [93] Concepción Burgos, María L. Campanario, David de la Peña, Juan A. Lara, David Lizcano, María A. Martínez (2018). Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout, *Computers & Electrical Engineering*, Volume 66, 2018, Pages 541-556, ISSN 0045-7906, <https://doi.org/10.1016/j.compeleceng.2017.03.005>.
- [94] Hashmia Hamsa, Simi Indiradevi, Jubilant J. Kizhakkethottam (2016). Student Academic Performance Prediction Model Using Decision Tree and Fuzzy Genetic Algorithm, *Procedia Technology*, Volume 25, 2016, Pages 326-332, ISSN 2212-0173, <https://doi.org/10.1016/j.protcy.2016.08.114>.
- [95] Dongho Kim, Yeonjeong Park, Meehyun Yoon, Il-Hyun Jo (2016). Toward evidence-based learning analytics: Using proxy variables to improve asynchronous online discussion environments, *The Internet and Higher Education*, Volume 30, 2016, Pages 30-43, ISSN 1096-7516, <https://doi.org/10.1016/j.iheduc.2016.03.002>.
- [96] Sumyea Helal, Jiuyong Li, Lin Liu, Esmaeil Ebrahimie, Shane Dawson, Duncan J. Murray, Qi Long (2018). Predicting academic performance by considering student heterogeneity, *Knowledge-Based Systems*, Volume 161, 2018, Pages 134-146, ISSN 0950-7051, <https://doi.org/10.1016/j.knosys.2018.07.042>.
- [97] Augusto Sandoval, Carlos Gonzalez, Rosa Alarcon, Karim Pichara, Maximiliano Montenegro (2018). Centralized student performance prediction in large courses based on low-cost variables in an institutional context, *The Internet and Higher Education*, Volume 37, 2018, Pages 76-89, ISSN 1096-7516, <https://doi.org/10.1016/j.iheduc.2018.02.002>.
- [98] Giora Alexandron, José A. Ruipérez-Valiente, Zhongzhou Chen, Pedro J. Muñoz-Merino, David E. Pritchard (2017). Copying@Scale: Using Harvesting Accounts for Collecting Correct Answers in a MOOC, *Computers & Education*, Volume 108, 2017, Pages 96-114, ISSN 0360-1315, <https://doi.org/10.1016/j.compedu.2017.01.015>.
- [99] Tony Cheng-Kui Huang, Chih-Hong Huang, Yung-Ting Chuang (2016). Change discovery of learning performance in dynamic educational environments, *Telematics and Informatics*, Volume 33, Issue 3, 2016, Pages 773-792, ISSN 0736-5853, <https://doi.org/10.1016/j.tele.2015.10.005>.
- [100] Elena Gaudio, Miguel Montero, Luis Talavera, Felix Hernandez-del-Olmo (2009). Supporting teachers in collaborative student modeling: A framework and an implementation, *Expert Systems with Applications*, Volume 36, Issue 2, Part 1, 2009, Pages 2260-2265, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2007.12.035>.
- [101] Ángel Serrano-Laguna, Javier Torrente, Pablo Moreno-Ger, Baltasar Fernández-Manjón (2014). Application of Learning Analytics in educational videogames, *Entertainment Computing*, Volume 5, Issue 4, 2014, Pages 313-322, ISSN 1875-9521, <https://doi.org/10.1016/j.entcom.2014.02.003>.
- [102] Young-Jin Lee (2012). Developing an efficient computational method that estimates the ability of students in a Web-based learning environment, *Computers & Education*, Volume 58, Issue 1, 2012, Pages 579-589, ISSN 0360-1315, <https://doi.org/10.1016/j.compedu.2011.09.008>.
- [103] Philippe Fournier-Viger, Usef Faghihi, Roger Nkambou, Engelbert Mephu Nguifo (2012). CMRules: Mining sequential rules common to several sequences, *Knowledge-Based Systems*, Volume 25, Issue 1, 2012, Pages 63-76, ISSN 0950-7051, <https://doi.org/10.1016/j.knosys.2011.07.005>.
- [104] Zailani Abdullah, Tutut Herawan, Noraziah Ahmad, Mustafa Mat Deris (2011). Mining significant association rules from educational data using critical relative support approach, *Procedia - Social and Behavioral Sciences*, Volume 28, 2011, Pages 97-101, ISSN 1877-0428, <https://doi.org/10.1016/j.sbspro.2011.11.020>.
- [105] Fabio Aiolli, Kerstin Bunte, Romain Hérault, Mikhail Kanevski (2016). Special issue: *Advances in artificial neural networks, machine learning and computational intelligence* Selected papers from the 23rd European Symposium on Artificial Neural Networks (ESANN 2015), *Neurocomputing*, Volume 192, 2016, Pages 1-2, ISSN 0925-2312, <https://doi.org/10.1016/j.neucom.2016.02.005>.

- [106] Sergio Salmeron-Majadas, Olga C. Santos, Jesus G. Boticario (2014). An Evaluation of Mouse and Keyboard Interaction Indicators towards Non-intrusive and Low Cost Affective Modeling in an Educational Context, *Procedia Computer Science*, Volume 35, 2014, Pages 691-700, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2014.08.151>.
- [107] Nurbiha A. Shukor, Zaidatun Tasir, Henny Van der Meijden (2015). An Examination of Online Learning Effectiveness Using Data Mining, *Procedia - Social and Behavioral Sciences*, Volume 172, 2015, Pages 555-562, ISSN 1877-0428, <https://doi.org/10.1016/j.sbspro.2015.01.402>.
- [108] Beatriz Fernández-Gallego, Manuel Lama, Juan C. Vidal, Manuel Mucientes (2013). Learning Analytics Framework for Educational Virtual Worlds, *Procedia Computer Science*, Volume 25, 2013, Pages 443-447, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2013.11.056>.
- [109] Chee Kian Leong, Yew Haur Lee, Wai Keong Mak (2012). Mining sentiments in SMS texts for teaching evaluation, *Expert Systems with Applications*, Volume 39, Issue 3, 2012, Pages 2584-2589, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2011.08.113>.
- [110] Setsuya Kurahashi, Keisuke Kuniyoshi, Effects of Collaborative Learning on a Complex Doubly Structured Network, *Procedia Computer Science*, Volume 35, 2014, Pages 910-917, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2014.08.165>.
- [111] Stefan Svetsky, Oliver Moravcik (2013). The Supportive System for the Processing of Human and Non-human Knowledge Sources and Combining Mining Techniques, *AASRI Procedia*, Volume 4, 2013, Pages 182-189, ISSN 2212-6716, <https://doi.org/10.1016/j.aasri.2013.10.028>.
- [112] Jorge Miguel, Santi Caballé, Fatos Xhafa (2017). Chapter 5 - Massive data processing for effective trustworthiness modeling, Chapter 7 - Trustworthiness in action: Data collection, processing, and visualization methods for real online courses, Editor(s): Jorge Miguel, Santi Caballé, Fatos Xhafa, In *Intelligent Data-Centric Systems, Intelligent Data Analysis for e-Learning*, Academic Press, 2017, Pages 83-104, ISBN 9780128045350, <https://doi.org/10.1016/B978-0-12-804535-0.00005-8>.
- [113] Zhi Liu, Sanya Liu, Lin Liu, Jianwen Sun, Xian Peng, Tai Wang (2016). Sentiment recognition of online course reviews using multi-swarm optimization-based selected features, *Neurocomputing*, Volume 185, 2016, Pages 11-20, ISSN 0925-2312, <https://doi.org/10.1016/j.neucom.2015.12.036>.
- [114] Hiroshi Ueda, Motonori Nakamura (2017). Data Analysis for Evaluation on Course Design and Improvement of “Cyberethics” Moodle Online Courses, *Procedia Computer Science*, Volume 112, 2017, Pages 2345-2353, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2017.08.204>.
- [115] Pavel Michlík, Mária Bielíková (2010). Exercises recommending for limited time learning, *Procedia Computer Science*, Volume 1, Issue 2, 2010, Pages 2821-2828, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2010.08.007>.
- [116] Miroslav Minović, Miloš Milovanović, Uroš Šošević, Miguel Ángel Conde González (2015). Visualisation of student learning model in serious games, *Computers in Human Behavior*, Volume 47, 2015, Pages 98-107, ISSN 0747-5632, <https://doi.org/10.1016/j.chb.2014.09.005>.
- [117] Gökhan Engin, Burak Aksoyer, Melike Avdagic, Damla Bozanlı, Umutcan Hanay, Deniz Maden, Gurdal Ertek (2014). Rule-based Expert Systems for Supporting University Students, *Procedia Computer Science*, Volume 31, 2014, Pages 22-31, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2014.05.241>.
- [118] Xiao Zhong, Zeyi Sun, Haoyi Xiong, Neil Heffernan, Md. Monirul Islam (2017). Learning Curve Analysis Using Intensive Longitudinal and Cluster-Correlated Data, *Procedia Computer Science*, Volume 114, 2017, Pages 250-257, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2017.09.035>.
- [119] Elena Gaudioso, Miguel Montero, Felix Hernandez-del-Olmo (2012). Supporting teachers in adaptive educational systems through predictive models: A proof of concept, *Expert Systems with Applications*, Volume 39, Issue 1, 2012, Pages 621-625, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2011.07.052>.
- [120] Muamer N. Mohammad, Norrozila Sulaiman, Osama Abdulkarim Muhsin (2011). A novel intrusion detection system by using intelligent data mining in weka environment, *Procedia Computer Science*, Volume 3, 2011, Pages 1237-1242, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2010.12.198>.

- [121] Roberto Legaspi, Raymund Sison, Kenichi Fukui, Masayuki Numao (2008). Cluster-based predictive modeling to improve pedagogic reasoning, *Computers in Human Behavior*, Volume 24, Issue 2, 2008, Pages 153-172, ISSN 0747-5632, <https://doi.org/10.1016/j.chb.2007.01.007>.
- [122] Sfrent, A and Pop, F. (2015). *Asymptotic scheduling for many task computing in Big Data platforms*. Information Sciences 319, pp.71-91. DOI: 10.1016/j.ins.2015.03.053
- [123] Vrejoiu, MH; Zamfir, MC and Florian, V. (2017). *Visualization of Big and Visual analytics Data. Approaches and trends*. 2017 | Romanian journal of information technology and automatic control-Revista romana de informatica si automatica 27 (2), pp.5-16. ISSN: 1220-1758
- [124] Jaba, E; Botezat, A and Balan, C (2010). *The analysis of the creativity determinants - case study on Romanian and Italian students*. 6th International Seminar on the Quality Management in Higher Education. Quality management in higher education, VOL 1, pp.443-446. ISBN978-973-662-567-1
- [125] Danubianu, M; Pentiuc, SG and Tobolcea, I. (2011). *Mining Association Rules Inside a Relational Database - A Case Study*. 2011 | Sixth International Multi-Conference on Computing in The Global Information Technology (ICCGI 2011), pp.14-19. ISBN978-1-61208-139-7
- [126] Dragos, S and Dragos, R. (2009). *WATEC: A Web Analytics Tool for Educational Content*. KEPT 2009: Knowledge Engineering Principles and Techniques, pp.320-+. ISSN2067-1180
- [127] Scutelnicu, LA; Bibiri, AD and Cristea, D (2015). *An approach for interconnecting lexical resources*. Proceedings of the 11TH International Conference 'Linguistic Resources and Tools for Processing the Romanian Language', 2015, pp.31-38
- [128] Topirceanu, A; Tiselice, D and Udrescu, M (2015). *The Fingerprint of Educational Platforms in Social Media: A Topological Study Using Online Ego-Networks*. International Conference on Social Media in Academia - Research and Teaching (SMART), pp.355-360
- [129] Titu, MA and Bucur, V. (2010). *The Analysis of the Managers' Behaviour in the Knowledge-Based Organizations from the Educational System*. International Multiconference of Engineers and Computer Scientists (IMECS 2010), VOLS I-III, pp.1789
- [130] Stan, OM. Assessing Corporate Induction by e-Learning: Ocular centrism and Impact. 14th International Scientific Conference on eLearning and Software for Education - eLearning Challenges and New Horizons. DOI10.12753/2066-026X-18-248
- [131] Benta, D; Bologa, G; (...); Dzitac, I. (2015). University Level Learning and Teaching via E-Learning Platforms. 3RD International Conference on Information Technology and Quantitative Management, ITQM 2015 55, pp.1366-1373. DOI10.1016/j.procs.2015.07.123
- [132] A. Stan, P. Bell and S. King (2012). *A grapheme-based method for automatic alignment of speech and text data*. IEEE Spoken Language Technology Workshop (SLT), 2012, pp. 286-290, doi: 10.1109/SLT.2012.6424237.
- [133] M. Danubianu, S. -G. Pentiuc, O. A. Schipor, M. Nestor and I. Ungureanu (2008). *Distributed Intelligent System for Personalized Therapy of Speech Disorders*. 2008 The Third International Multi-Conference on Computing in the Global Information Technology (ICCGI 2008), 2008, pp. 166-170, doi: 10.1109/ICCGI.2008.31.
- [134] H. Cucu, L. Besacier, C. Burileanu and A. Buzo (2011). *Investigating the role of machine translated text in ASR domain adaptation: Unsupervised and semi-supervised methods*. 2011 IEEE Workshop on Automatic Speech Recognition & Understanding, 2011, pp. 260-265, doi: 10.1109/ASRU.2011.6163941.
- [135] A. Caranica, H. Cucu, C. Burileanu, F. Portet and M. Vacher (2017). *Speech recognition results for voice-controlled assistive applications*. 2017 International Conference on Speech Technology and Human-Computer Dialogue (SpeD), 2017, pp. 1-8, doi: 10.1109/SPED.2017.7990438.
- [136] D. Andone, S. Vert, M. Frydenberg and R. Vasii (2018). *Open Virtual Reality Project to Improve Students' Skills*. 2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT), 2018, pp. 6-10, doi: 10.1109/ICALT.2018.00008.

- [137] S. M. Feraru and M. Dan Zbancioc (2013). *Emotion recognition in Romanian language using LPC features*. 2013 E-Health and Bioengineering Conference (EHB), 2013, pp. 1-4, doi: 10.1109/EHB.2013.6707314.
- [138] A. Friesel, D. Cojocaru and K. Avramides (2015). *Identifying how PELARS-project can support the development of new curriculum structures in engineering education*. 2015 3rd Experiment International Conference (exp.at'15), 2015, pp. 219-223, doi: 10.1109/EXPAT.2015.7463269.
- [139] I. Pavaloi, E. Musca and F. Rotaru (2013). *Emotion recognition in audio records*. International Symposium on Signals, Circuits and Systems ISSCS2013, 2013, pp. 1-4, doi: 10.1109/ISSCS.2013.6651236.
- [140] C. Savulescu, Z. Polkowski, D. I. Cosmin and B. C. Elena (2015). *Security in e-learning systems*. 2015 7th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), 2015, pp. WE-19-WE-24, doi: 10.1109/ECAI.2015.7301225.
- [141] V. Spiridon and E. Slusanschi (2013). *N-Body Simulations with GADGET-2*. 2013 15th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, 2013, pp. 526-533, doi: 10.1109/SYNASC.2013.75.
- [142] V. Maria and F. Mone (2018). *Big data services based on mobile data and their strategic importance*. 2018 7th International Conference on Computers Communications and Control (ICCCC), 2018, pp. 276-281, doi: 10.1109/ICCCC.2018.8390471.
- [143] M. Colhon, M. Cerban, A. Becheru and M. Teodorescu (2016). *Polarity shifting for Romanian sentiment classification*. 2016 International Symposium on INnovations in Intelligent SysTems and Applications (INISTA), 2016, pp. 1-6, doi: 10.1109/INISTA.2016.7571834.
- [144] M. Lazăr and D. Militaru (2013). *A Romanian language modeling using linguistic factors*. 2013 7th Conference on Speech Technology and Human - Computer Dialogue (SpeD), 2013, pp. 1-6, doi: 10.1109/SpeD.2013.6682649.
- [145] S. Brad, M. Fulea, B. Mocan, A. Duca and E. Brad (2008). *Software platform for supporting open innovation*. 2008 IEEE International Conference on Automation, Quality and Testing, Robotics, 2008, pp. 224-229, doi: 10.1109/AQTR.2008.4588916.
- [146] A. Bîncă, G. Muraru and V. Posea (2014). *IT skills in Romania*. 2014 RoEduNet Conference 13th Edition: Networking in Education and Research Joint Event RENAM 8th Conference, 2014, pp. 1-5, doi: 10.1109/RoEduNet-RENAM.2014.6955320.
- [147] M. Orzan, G. Orzan, G. Spiridon and T. Neacsu (2010). *Informatic online system for dissemination of scientific research results for intellectual property rights*. 2010 IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR), 2010, pp. 1-6, doi: 10.1109/AQTR.2010.5520782.
- [148] S. Moica, C. Veres Harea and L. Marian (2018). *Effects of Suggestion System on Continuous Improvement: A Case Study*. 2018 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), 2018, pp. 592-596, doi: 10.1109/IEEM.2018.8607804.
- [149] D. Trandabăț (2011). *Mining Romanian texts for semantic knowledge*. 2011 11th International Conference on Intelligent Systems Design and Applications, 2011, pp. 1062-1066, doi: 10.1109/ISDA.2011.6121799.
- [150] S. Matei, C. Rughinis and R. Rughinis (2017). *Big Data, Old Users, Personal Worlds: A Survey of Challenges and Resistance to Big Data Analytics in the EU*. 2017 21st International Conference on Control Systems and Computer Science (CSCS), 2017, pp. 175-181, doi: 10.1109/CSCS.2017.31.
- [151] B. Patrut, S. E. Vârlan and T. Socaciu (2008). *ContTest – A Multi-agent System for Accounting Education*. 2008 The Third International Multi-Conference on Computing in the Global Information Technology (ICCGI 2008), 2008, pp. 217-222, doi: 10.1109/ICCGI.2008.17.
- [152] D. Cojocaru, R. T. Tanasie and A. Friesel (2017). *Updating Engineering Education Technologies: UCV in PELARS*. 2017 27th EAAEIE Annual Conference (EAAEIE), 2017, pp. 1-6, doi: 10.1109/EAAEIE.2017.8768630.
- [153] R. Žitný, T. Szabó, I. Pšenáková, Z. Illés and V. H. Bakonyi (2016). *Using mobile technologies in university education*. 2016 International Conference on Emerging eLearning Technologies and Applications (ICETA), 2016, pp. 387-392, doi: 10.1109/ICETA.2016.7802042.

- [154] M. Zaharia and D. Enachescu (2013). *Considerations on the evolution of attention paid to education in Romania compared to EU, in the 1998–2010 period*. 2013 36th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2013, pp. 547-552.
- [155] A. Anisie, M. Dinsoreanu, C. Lemnar and R. Potolea (2014). *Extracting opinion holders and targets in Romanian texts*. 2014 IEEE 10th International Conference on Intelligent Computer Communication and Processing (ICCP), 2014, pp. 37-42, doi: 10.1109/ICCP.2014.6936977.
- [156] V. Holicov (2010). *Changes and innovations in the learning process for the I-society*. 2010 International Conference on Information Society, 2010, pp. 471-473, doi: 10.1109/i-Society16502.2010.6018751.
- [157] Mircea Eremia; Chen-Ching Liu; Abdel-Aty Edris (2016). *Decision Trees in Advanced Solutions in Power Systems: HVDC, FACTS, and Artificial Intelligence*, IEEE, 2016, pp.819-844, doi: 10.1002/9781119175391.ch18.
- [158] D. Cojocaru, D. Popescu, R. T. Tănasie and T. Ward (2014). *Programme policies in European engineering higher education system*. 2014 Information Technology Based Higher Education and Training (ITHET), 2014, pp. 1-6, doi: 10.1109/ITHET.2014.7155698.
- [159] A. Lazăr and C. Langa (2016). *Related review of theories developed on using CLIL in primary schools in Romania*. 2016 8th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), 2016, pp. 1-5, doi: 10.1109/ECAI.2016.7861135.
- [160] E. M. Alborno and D. H. Milone (2017). Emotion recognition in never-seen languages using a novel ensemble method with emotion profiles. *IEEE Transactions on Affective Computing*, vol. 8, no. 1, pp. 43-53, 1 Jan.-March 2017, doi: 10.1109/TAFFC.2015.2503757.
- [161] D. Cmur, T. Mutapcic and I. Bilic (2009). MIRACLE—Model for Integration of Remote Laboratories in Courses that Use Laboratory and e-Learning Systems in *IEEE Transactions on Learning Technologies*, vol. 2, no. 4, pp. 275-288, Oct.-Dec. 2009, doi: 10.1109/TLT.2009.48.
- [162] A. Ucan, B. Naderalvojud, E. A. Sezer and H. Sever (2016). *SentiWordNet for New Language: Automatic Translation Approach*. 2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), 2016, pp. 308-315, doi: 10.1109/SITIS.2016.57.
- [163] Á. Beszédes and L. Vidács (2016). *Academic and Industrial Software Testing Conferences: Survey and Synergies*. 2016 IEEE Ninth International Conference on Software Testing, Verification and Validation Workshops (ICSTW), 2016, pp. 240-249, doi: 10.1109/ICSTW.2016.30.
- [164] D. Bamman, M. Carney, J. Gillick, C. Hennesy and V. Sridhar (2017). *Estimating the Date of First Publication in a Large-Scale Digital Library*. 2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL), 2017, pp. 1-10, doi: 10.1109/JCDL.2017.7991569.
- [165] M. Kersten and T. O'Brien (2011). *A Cross-Cultural Collaborative Learning Project: Learning about Cultures and Research Methods*. 2011 44th Hawaii International Conference on System Sciences, 2011, pp. 1-9, doi: 10.1109/HICSS.2011.6.
- [166] B. Grobowski and V. Vidacek-Hainš (2015). *The relationship between cultural dimensions and the quality of communication in education*. 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2015, pp. 652-656, doi: 10.1109/MIPRO.2015.7160353.
- [167] D. Pal'ová, N. M. Novak and V. Weidinger (2017). *Digital learning as a tool to overcome school failure in minority groups*. 2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2017, pp. 767-772, doi: 10.23919/MIPRO.2017.7973525.
- [168] Santos, O. C. (Ed.). (2011). *Educational Recommender Systems and Technologies: Practices and Challenges: Practices and Challenges. Chapter 7. Clustering of the Web Search Results in Educational Recommender System*. IGI Global, 2011. ISBN161 350490X, 9781613504901, pp. 154–181.
- [169] P. Ș. Popescu, M. Mocanu, D. Dan Burdescu and M. C. Mihăescu (2015). *Messaging activity impact on learners profiling*. 2015 6th International Conference on Information, Intelligence, Systems and Applications (IISA), 2015, pp. 1-6, doi: 10.1109/IISA.2015.7387980.

- [170] Albeanu, G., Madsen, H., & Popentiu-Vladicescu, F. (2016, April). Learning from nature: Nature-inspired algorithms. In *The International Scientific Conference eLearning and Software for Education* (Vol. 2, p. 477). "Carol I" National Defence University.
- [171] Albeanu, G., Țarcă, R. C., Popențiu-Vlădicescu, F., & Pașc, I. M. (2010). Interoperability assurance for remote mechatronic laboratories used for virtual training. *ELearning & Software for Education*.
- [172] Maria Goga, Shade Kuyoro, Nicolae Goga (2015). A Recommender for Improving the Student Academic Performance. *Procedia - Social and Behavioral Sciences*, Volume 180, 2015, Pages 1481-1488, ISSN 1877-0428, <https://doi.org/10.1016/j.sbspro.2015.02.296>.
- [173] Nicolae Nistor, Ángel Hernández-García (2018). What types of data are used in learning analytics? An overview of six cases. *Computers in Human Behavior*, Volume 89, 2018, Pages 335-338, ISSN 0747-5632, <https://doi.org/10.1016/j.chb.2018.07.038>.
- [174] Elena Susnea (2018). How Big Data analytics will reshape e-learning? *ELearning challenges and new horizons*, VOL 4, Pp. 72-76. București, Meeting 14th International Scientific Conference on eLearning and Software for Education - eLearning Challenges and New Horizons. DOI10.12753/2066-026X-18-225
- [175] Dascalu, MI; Tesila, B; (...); Radu, IC. (2018). *A Big Data Analytics Tool for Social Learning Management Systems*. 14th International Scientific Conference on eLearning and Software for Education - eLearning Challenges and New Horizons. VOL 2, pp.152-159. DOI10.12753/2066-026X-18-09
- [176] Bucos, M and Dragulescu, B. (2018). *Predicting Student Success Using Data Generated in Traditional Educational Environments*. 2018 | TEM Journal-Technology Education Management Informatics 7 (3), pp.617-625. DOI 10.18421/TEM73-19
- [177] Mihaescu, C; Burdescu, DD; (...); Ionascu, C. (2009). *Obtaining Knowledge Using Educational Data Mining*. 3rd International Symposium on Intelligent and Distributed Computing (IDC 2009). *Intelligent Distributed Computing III* volum 237, pp.257+
- [178] Danubianu, Mirela. (2015). *A data preprocessing framework for students' outcome prediction by data mining techniques*. 19th International Conference on System Theory, Control and Computing (ICSTCC). Pp. 836-841. 10.1109/ICSTCC.2015.7321398.
- [179] Susnea, E (2013). Using Data Mining in e-learning a generic framework for military education. 9th International Conference eLearning and Software for Education. *Quality and efficiency in e-learning*, VOL 1, pp.411-415
- [180] Bara, A; Andreescu, A; (...); Velicanu, M (2013). Data mining solutions for determining student's profile. 9th International Conference eLearning and Software for Education. *Quality and efficiency in e-learning*, VOL 2, pp.284-289
- [181] Susnea, E (2011). *Data mining techniques used in on-line military training*. 7th International Scientific Conference eLearning and Software for Education. Anywhere, anytime - education on demand, VOL I, pp.201-205
- [182] Vilanova, R; Vicario, J; (...); Paganoni, A. (2017). *SPEET: an international collaborative experience in data mining for education*. 2017 | 10TH International Conference of Education, Research and Innovation (ICERI2017), pp.1648-1656. ISSN: 2340-1095
- [183] Oprea, C; Popescu, DM; (...); Barbu, I (2017). A data mining based model to improve university management. *Journal of Science and Arts* (2), pp.285-296. ISSN: 1844-9581
- [184] Bresfelean, VP; Bresfelean, M; (...); Comes, CA (2008). *Determining students' academic failure profile founded on data mining methods*. 30th International Conference on Information Technology Interfaces, pp.317-+. DOI10.1109/ITI.2008.4588429
- [185] Susnea, E. (2009). *Classification techniques used in Educational System*. ICVL 2009 - Proceedings of the 4th International Conference on Virtual Learning, pp.376-379

- [186] Danubianu, M (2016). *First Steps Towards Curriculum Development Decisions From Raw Educational Data*. 13TH International Conference on Development and Application Systems (DAS 2016), pp.281-286. DOI:10.1109/DAAS.2016.7492587
- [187] Crossley, S; Paquette, L; (...); Baker, RS (2016). *Combining Click-Stream Data with NLP Tools to Better Understand MOOC Completion*. LAK '16 Conference Proceedings: The Sixth International Learning Analytics & Knowledge Conference, pp.6-14. DOI10.1145/2883851.2883931.
- [188] Vilanova, R; Vicario, J; (...); Paganoni, A (2018). *SPEET: Software Tools for Academic Data Analysis*. EDULEARN18: 10TH International Conference on Education and New Learning Technologies, pp.7616-7625
- [189] Dragos, S; Halita, D and Sacarea, C (2015). *Attractors in Web Based Educational Systems a Conceptual Knowledge Processing Grounded Approach*. 8th International Conference on Knowledge Science, Engineering and Management (KSEM), KSEM 2015 9403, pp.190-195. DOI10.1007/978-3-319-25159-2_18
- [190] Dragos, S; Halita, D and Sacarea, C. (2015). *Behavioral pattern mining in web based educational systems*. 2015 23RD International Conference on Software, Telecommunications and Computer Networks (SOFTCOM), pp.215-219. ISBN978-1-5090-0053-1
- [191] Dragos, SM; Halita, DF and Sacarea, C (2016). *Distilling Conceptual Structures from Weblog Data Using Polyadic FCA*. 22nd International Conference on Conceptual Structures (ICCS). Graph-Based Representation and Reasoning (ICCS 2016) 9717, pp.151-159. DOI10.1007/978-3-319-40985-6_12
- [192] Mihaescu, MC; Tanasie, AV; (...); Trausan-Matu, S. (2016). *Extracting Patterns from Educational Traces via Clustering and Associated Quality Metrics*. 17th International Conference on Artificial Intelligence - Methodology, Systems and Applications (AIMSA). 2016 9883, pp.109-118. DOI10.1007/978-3-319-44748-3_11
- [193] Bresfelean, VP; Mocean, L and Hajdu-Macelararu, M (2016). *Educational portals. Current studies and steps for creating the etrajjectory portal*. 3rd International Multidisciplinary Scientific Conference on Social Sciences and Arts, SGEM 2016, BK 1: Psychology and psychiatry, sociology and healthcare, education conference proceedings, VOL III, pp.271-278. ISBN978-619-7105-72-8.
- [194] Pupezescu, V (2016). *Distributed neural structures in adaptive eLearning systems*. Proceedings of the 11th International Conference on Virtual Learning, pp.302-309
- [195] Bresfelean, VP; Ghisoiu, N; (...); Sitar-Taut, DA (2009). *Towards the Development of Decision Support in Academic Environments*. Proceedings of the ITI 2009 31st International Conference on Information Technology Interfaces, pp.343-348. DOI10.1109/ITI.2009.5196106
- [196] Leon, F and Popescu, E (2017). *Using Large Margin Nearest Neighbor Regression Algorithm to Predict Student Grades Based on Social Media Traces*. 7th International Conference in Methodologies and Intelligent Systems for Technology Enhanced Learning (MIS4TEL). 617, pp.12-19. DOI10.1007/978-3-319-60819-8_2
- [197] Burdescu, DD (2013). TESYS- an e-learning system. 9th International Conference eLearning and Software for Education. *Quality And Efficiency In E-Learning*, VOL 3, pp.255-260
- [198] Constantinov, C; Popescu, PS; (...); Mocanu, ML (2015). *Preliminary results of a curriculum adjuster based on professional network analysis*. 19th International Conference on System Theory, Control and Computing (ICSTCC), pp. 860-865
- [199] Popescu, PS; Mocanu, M; (...); Mihaescu, MC (2016). *Design of an Advanced Smart Forum for Tesys e-Learning Platform*. 12th IFIP WG 12.5 International Conference on Artificial Intelligence Applications and Innovations (AIAI), 475, pp.305-316. DOI10.1007/978-3-319-44944-9_26
- [200] Susnea, E (2013). *Improving Decision Making Process in Universities: A Conceptual Model of Intelligent Decision Support System*. 5th International Conference EDU-WORLD - Education Facing Contemporary World Issues, 76, pp.795-800. DOI10.1016/j.sbspro.2013.04.208
- [201] Topirceanu, A and Grosseck, G (2017). *Decision tree learning used for the classification of student archetypes in online courses*. 21st International Conference on Knowledge - Based and Intelligent Information and Engineering Systems (KES), 112, pp.51-60. DOI10.1016/j.procs.2017.08.021

- [202] C. Constantinov, P. Ș. Popescu, C. M. Poteraș and M. L. Mocanu (2015). *Preliminary results of a curriculum adjuster based on professional network analysis*. 19th International Conference on System Theory, Control and Computing (ICSTCC), 2015, pp. 860-865, doi: 10.1109/ICSTCC.2015.7321402.
- [203] F. Trif, C. Lemnaru and R. Potolea, "Identifying the user typology for adaptive e-learning systems," 2010 IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR), 2010, pp. 1-6, doi: 10.1109/AQTR.2010.5520728.
- [204] T. Gordiyenko, O. Velychko and A. Salceanu (2018). *The Group Expert Evaluation in Electrical Engineering Education*. 2018 International Conference and Exposition on Electrical and Power Engineering (EPE), 2018, pp. 0021-0026, doi: 10.1109/ICEPE.2018.8559787.
- [205] C. M. Mihăescu, O. M. Teodorescu, P. Ș. Popescu and M. L. Mocanu (2017). *Learning analytics solution for building personalized quiz sessions*. 2017 18th International Carpathian Control Conference (ICCC), 2017, pp. 140-145, doi: 10.1109/CarpathianCC.2017.7970386.
- [206] Ianculescu, M; Alexandru, A and Gheorghe-Moisii M. (2017). *Harnessing the Potential of Big Data in Romanian Healthcare*. 5th International Symposium on Electrical and Electronics Engineering (ISEEE). ISSN 2378-3907
- [207] Geman, O; Chiuchisan, I; (...); Milici, LD (2018). Deep Learning Tools for Human Microbiome Big Data. Soft Computing Applications, SOFA 2016, VOL 1. Book Series Advances in *Intelligent Systems and Computing*, Volume 633, Pp. 265-275. DOI 10.1007/978-3-319-62521-8_21
- [208] Nafornta, C; Isar, A; (...); Caba, C (2017). *Solution of a Big Data Problem. Simulator for Denoising of Single Look Complex SAR Images*. 2017 International Symposium on Signals, Circuits and Systems (ISSCS). DOI 10.1109/ISSCS.2017.8034865
- [209] Chee Kian Leong, Yew Haur Lee, Wai Keong Mak (2012). Mining sentiments in SMS texts for teaching evaluation. *Expert Systems with Applications*, Volume 39, Issue 3, 2012, Pages 2584-2589, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2011.08.113>.
- [210] Elgendy N., Elragal A. (2014). Big Data Analytics: A Literature Review Paper. In: Perner P. (eds) *Advances in Data Mining. Applications and Theoretical Aspects*. ICDM 2014. Lecture Notes in Computer Science, vol 8557. Springer, Cham. https://doi.org/10.1007/978-3-319-08976-8_16.
- [211] Patgiri, Ripon & Ahmed, Arif. (2016). *Big Data: The V's of the Game Changer Paradigm*. Conference: 18th IEEE High Performance Computing and Communications. 10.1109/HPCC-SmartCity-DSS.2016.0014.
- [212] Kaynak, Okay & Yin, Shen. (2015). *Big Data for Modern Industry: Challenges and Trends [Point of View]*. Proceedings of the IEEE. 103. 143-146. 10.1109/JPROC.2015.2388958.
- [213] Mirela Danubianu, Adina Bărilă. *Big Data vs. Data Mining for Social Media Analytics*. Social Media in Academia: Research and Teaching, Timișoara, 18-21 septembrie 2014, pag. 261-270.
- [214] Marcu, Daniela; Danubianu, Mirela; Pentiuc, Gheorghe. *Big Data Technology - valuable tool for healthcare research*. International Journal of Computer Science and Network Security, April 30, 2019. ISSN: 1738-7906. Vol.19, No.4
- [215] Kapil, Gayatri & Agrawal, Alka & Khan, Prof. Raees. (2016). *A study of big data characteristics*. 1-4. 10.1109/CESYS.2016.7889917.
- [216] Hussien, Abou-El-Ela. (2020). *Fifty-Six Big Data V's Characteristics and Proposed Strategies to Overcome Security and Privacy Challenges (BD2)*. Journal of Information Security. Vol. 11. 304-328. 10.4236/jis.2020.114019.
- [217] MarketsandMarkets. *Big Data Market by Component, Deployment Mode, Organization Size, Business Function (Operations, Finance, and Marketing and Sales), Industry Vertical (BFSI, Manufacturing, and Healthcare and Life Sciences), and Region - Global Forecast to 2025*. Online: <https://www.marketsandmarkets.com/Market-Reports/big-data-market-1068.html>.

- [218] Y. Demchenko, C. de Laat and P. Membrey (2014). *Defining architecture components of the Big Data Ecosystem*. 2014 International Conference on Collaboration Technologies and Systems (CTS), Minneapolis, MN, USA, 2014, pp. 104-112, doi: 10.1109/CTS.2014.6867550.
- [219] Apache Hadoop – documentație oficială. Online: <https://hadoop.apache.org/>.
- [220] Alexandru, A., & Coardos, D. (2017). *Big Data – concepte, arhitecturi și tehnologii*, Art. 02 – Vol. 27 – Nr. 1 – 2017.
- [221] HDFS Architecture Guide – Data Organization. Online: https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html#Data+Organization.
- [222] K. Shvachko, H. Kuang, S. Radia and R. Chansler (2010). *The Hadoop Distributed File System*. 2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST), Incline Village, NV, USA, 2010, pp. 1-10, doi: 10.1109/MSST.2010.5496972.
- [223] Microsoft. Big Data Architecture. Documentație oficială. Online: <https://docs.microsoft.com/en-us/azure/architecture/data-guide/big-data/>.
- [224] Yang, Peilin & Thiagarajan, Srikanth & Lin, Jimmy. (2018). *Robust, Scalable, Real-Time Event Time Series Aggregation at Twitter*. SIGMOD '18: Proceedings of the 2018 International Conference on Management of Data. 595-599. 10.1145/3183713.3190663.
- [225] Zschörnig, Theo & Wehlitz, Robert & Franczyk, Bogdan. (2017). *A Personal Analytics Platform for the Internet of Things - Implementing Kappa Architecture with Microservice-based Stream Processing*. 733-738. 10.5220/0006355407330738.
- [226] AWS Outposts. Online: <https://aws.amazon.com/outposts/>.
- [227] Manekar, Amitkumar & Gera, Pradeepini. (2018). *Studying Cloud as IaaS for Big Data Analytics: Opportunity, Challenges*. International Journal of Engineering & Technology. 7. 909. 10.14419/ijet.v7i2.7.11094.
- [228] Magalhaes, Guilherme & Roloff, Eduardo & Maillard, Nicolas. (2012). *Developing on Google App Engine*.
- [229] W. Fang, B. Jin, B. Zhang, Y. Yang and Z. Qin (2011). *Design and evaluation of a Pub/Sub service in the cloud*. 2011 International Conference on Cloud and Service Computing, 2011, pp. 32-39, doi: 10.1109/CSC.2011.6138542.
- [230] Syed, Hassan & Gani, Abdullah & Ahmad, Raja & Khan, Khurram & Ahmed, Abdelmutilib Ibrahim Abdalla. (2017). *Cloud Monitoring: A Review, Taxonomy, and Open Research Issues*. Journal of Network and Computer Applications. 98. 10.1016/j.jnca.2017.08.021.
- [231] *Big data – the cloud way*. Google Cloud Platform Blog. Online: <https://cloudplatform.googleblog.com/2015/04/big-data-cloud-way.html>
- [232] Google Cloud for AWS Professionals. Online: <https://cloud.google.com/docs/compare/aws>.
- [233] Intellipaat – furnizor global de cursuri de formare. Online: <https://intellipaat.com/blog/aws-vs-google-cloud/#no5>.
- [234] El Morr, Christo & Ali-Hassan, Hossam. (2019). *Descriptive, Predictive, and Prescriptive Analytics: A Practical Introduction*. 10.1007/978-3-030-04506-7_3.
- [235] Katerina Lepenioti, Alexandros Bousdekis, Dimitris Apostolou, Gregoris Mentzas (2020). Prescriptive analytics: Literature review and research challenges. *International Journal of Information Management*, Volume 50, 2020, Pages 57-70, ISSN 0268-4012, <https://doi.org/10.1016/j.ijinfomgt.2019.04.003>.
- [236] D. Marcu and M. Danubianu (2020). *Sentiment Analysis from Students' Feedback: A Romanian High School Case Study*. 2020 International Conference on Development and Application Systems (DAS), Suceava, Romania, 2020, pp. 204-209, doi: 10.1109/DAS49615.2020.9108927.
- [237] Ahlawat, Khyati & Singh, Amit. (2016). *Domain wise Analytics of Big Data- a study*. 7th International Conference on Communication, Computing and Virtualization 2016. Published by Elsevier B.V.

- [238] Baig, M.I., Shuib, L. & Yadegaridehkordi, E. (2020). *Big data in education: a state of the art, limitations, and future research directions*. International Journal of Education and Technology in Higher Education 17, 44 (2020). <https://doi.org/10.1186/s41239-020-00223-0>.
- [239] Marcu, Daniela; Danubianu, Mirela. (2019). *Learning Analytics or Educational Data Mining? This is the Question....* BRAIN. Broad Research in Artificial Intelligence and Neuroscience, [S.l.], v. 10, p. 1-14, oct. 2019. ISSN 2067-3957.
- [240] Baker RSJd, Gowda SM, Corbett AT. (2011). *Automatically detecting a student's preparation for future learning: help use is key*. Fourth International Conference on Educational Data Mining. Eindhoven, The Netherlands; 2011, 179–188.
- [241] Siemens, G. 2011. *About: Learning analytics and knowledge*. 1st International Conference on Learning Analytics and Knowledge, 2011.
- [242] Daniela Marcu, Mirela Danubianu (2021). *Sentiment Analysis From Images - Comparative Study of SAI-G and SAI-C Models' Performances Using AutoML Vision Service from Google Cloud and Clarifai Platform*. IJCSNS International Journal of Computer Science and Network Security, VOL.21 No.9, September 2021. DOI: <https://doi.org/10.22937/IJCSNS.2021.21.9.25>.
- [243] S. Athmaja, M. Hanumanthappa and V. Kavitha (2017). *A survey of machine learning algorithms for big data analytics*. 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, India, 2017, pp. 1-4, doi: 10.1109/ICIIECS.2017.8276028.
- [244] Lucian Busoni, Robert Babuska, Bart De Schutter, Damien Ernst (2010). *Reinforcement Learning and Dynamic Programming Using Function Approximators*. CRC Press, Automation and Control Engineering Series. April 2010, 280 pages, ISBN 978-1439821084.
- [245] Reinforcement Learning. GeeksforGeeks. Online: <https://www.geeksforgeeks.org/what-is-reinforcement-learning/>. Accesat la 26 aprilie 2021.
- [246] Flach, Peter (2012). *Machine Learning: The Art and Science of Algorithms that Make Sense of Data*. Cambridge University Press. ISBN 978-1-107-42222-3.
- [247] Chiru, C., Trausan-Matu, S, Rebedea, T. (2008). O îmbunătățire a performanțelor algoritmului KNN în sistemele de recomandare pe web. In: Buraga, S.C., Juvina, I. (Eds.) *Interacțiune Om-Calculator 2008*. ISSN 1843-4460, (Conferința Națională de Interacțiune Om-Calculator, Iași 18-19 Septembrie 2008), Editura MatrixROM București, pp.41-48.
- [248] Hu, L.Y., Huang, M.W., Ke, SW. et al. (2016). *The distance function effect on k-nearest neighbor classification for medical datasets*. SpringerPlus 5, 1304. <https://doi.org/10.1186/s40064-016-2941-7>
- [249] Sara Makki, Rafiqul Haque, Yehia Taher, Zainab Assaghir, Mohand-Said Hacid, et al. (2018). *A CostSensitive Cosine Similarity K-Nearest Neighbor for Credit Card Fraud Detection*. Big Data and Cyber-Security Intelligence, Dec 2018, Beirut, Lebanon. fihal-02353075f.
- [250] J. Nie, S. Yang, Q. Ren and D. Su (2018). *A Novel Classification Method Based on Adaboost for Electromagnetic Emission Characteristics*. 2018 International Applied Computational Electromagnetics Society Symposium - China (ACES), Beijing, China, 2018, pp. 1-3, doi: 10.23919/ACCESS.2018.8669248.
- [251] Gavin Hackelin (2017). *Mastering Machine Learning with scikit-learn* – second edition, 2017 Packt Publishing Ltd., Birmingham UK, p: 141-148, 88-92. ISBN 978-1-78829-987-9.
- [252] Zhu, Ji & Rosset, Saharon & Zou, Hui & Hastie, Trevor. (2006). *Multi-class AdaBoost. Statistics and its interface*. 2. 10.4310/SII.2009.v2.n3.a8.
- [253] D. Marcu, M. Danubianu, C. Simionescu (2021). *Comparative analysis of predictive models on online education in context of Covid-19 – a case study*. INTED2021 Proceedings, pp. 4403-4412.
- [254] R. Berwick, Village Idiot. SVMs: "A New. Generation of Learning Algorithms". Online: <https://web.mit.edu/6.034/wwwbob/svm.pdf>. Accesat la 4 aprilie 2021.

- [255] Souza, C. R. (2010). *Kernel functions for machine learning applications*. Creative Commons Attribution-Noncommercial-Share Alike, 3, 29.
- [256] Scikit Learn. *Naïve Bayes*.
Online: https://scikit-learn.org/stable/modules/naive_bayes.html. Accesat la 1 februarie 2023.
- [257] Liviu Ciortuz (2020). *Fundamente matematice ale învățării automate: Exerciții*. Suport de curs – Facultatea de Informatică, Iași. Online: <https://profs.info.uaic.ro/~ciortuz/ML.ex-book/Math4ML/book.Math4ML.26oct2020.pdf>. Accesat la 1 februarie 2023.
- [258] Rish, Irina. (2001). *An Empirical Study of the Naïve Bayes Classifier*. IJCAI 2001 Work Empir Methods Artif Intell. 3.
- [259] Clustering algorithms. *Suport de curs – Google*. Online: <https://developers.google.com/machine-learning/clustering/clustering-algorithms>. Accesat la 27 aprilie 2021
- [260] M. Emre Celebi, Hassan A. Kingravi, Patricio A. Vela (2013). A comparative study of efficient initialization methods for the k-means clustering algorithm. *Expert Systems with Applications*, Volume 40, Issue 1, 2013, Pages 200-210, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2012.07.021>. (<https://www.sciencedirect.com/science/article/pii/S0957417412008767>)
- [261] Ortega, Joaquín & Almanza-Ortega, Nelva & Vega-Villalobos, Andrea & Pazos-Rangel, Rodolfo & Zavala-Diaz, José Crispin & Martínez-Rebollar, Alicia. (2019). The K-Means Algorithm Evolution. 10.5772/intechopen.85447.
- [262] Santra, A. K., & Christy, C. J. (2012). *Genetic algorithm and confusion matrix for document clustering*. International Journal of Computer Science Issues (IJCSI), 9(1), 322.
- [263] HE Haibo, MA Yunqian (2013). *Imbalanced Learning: Foundations, Algorithms, and Applications* 1st Edition, Kindle Edition, Wilwy-IEEE Press, ASIN: B00DAB28YM, ISBN-13: 978-1118074626.
- [264] Keilwagen, Jens & Grosse, Ivo & Grau, Jan. (2014). *Area under Precision-Recall Curves for Weighted and Unweighted Data*. PloS one. 9. e92209. 10.1371/journal.pone.0092209.
- [265] Boyd, Kendrick & Eng, Kevin & Page, C. (2013). *Area under the Precision-Recall Curve: Point Estimates and Confidence Intervals*. 8190. 451-466. 10.1007/978-3-642-40994-3_29.
- [266] Documentație Scikit-Learn, *Precision-Recall*.
Online: https://scikit-learn.org/stable/auto_examples/model_selection/plot_precision_recall.html.
- [267] T. Young, D. Hazarika, S. Poria and E. Cambria (2018). *Recent Trends in Deep Learning Based Natural Language Processing [Review Article]*. IEEE Computational Intelligence Magazine, vol. 13, no. 3, pp. 55-75, Aug. 2018, doi: 10.1109/MCI.2018.2840738.
- [268] Khurana, Diksha & Koli, Aditya & Khatter, Kiran & Singh, Sukhdev (2017). *Natural Language Processing: State of The Art, Current Trends and Challenges*.
- [269] Thangaraj, M. & Sivakami, M. (2018). *Text Classification Techniques: A Literature Review*. Interdisciplinary Journal of Information, Knowledge, and Management. 13. 117-135. 10.28945/4066.
- [270] Thakor, P., & Sasi, S. (2015). *Ontology-based sentiment analysis process for social media content*. 2015 INNS Conference on Big Data. Procedia Computer Science, 53, 199-207.
- [271] Paltoglou, Georgios & Thelwall, Mike. (2010). *A Study of Information Retrieval Weighting Schemes for Sentiment Analysis*. ACL 2010 - 48th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference. 1386-1395.
- [272] Sidorov, Grigori & Castillo, Francisco & Stamatatos, Efstathios & Gelbukh, Alexander & Chanona-Hernández, Liliana. (2014). *Syntactic N-grams as machine learning features for natural language processing*. Expert Systems with Applications: An International Journal. 41. 853-860. 10.1016/j.eswa.2013.08.015.

- [273] Nicholls, Chris & Song, Fei. (2009). *Improving sentiment analysis with Part-of-Speech weighting*. Proceedings of the 2009 International Conference on Machine Learning and Cybernetics. 3. 1592 - 1597. 10.1109/ICMLC.2009.5212278.
- [274] Hutto, C.J. & Gilbert, Eric. (2015). *VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text*. Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014.
- [275] Turliuc, Dana & Cucu, Andrei & Costea, Claudia & Mohan, Aurel George & Ciurea, Alexandru. (2019). *Gheorghe Marinescu (1863-1938) and La Cellule Nerveuse. 110 years since writing of the major chapter of the Old Testament of neuropathology*. Revista medico-chirurgicala a Societatii de Medici si Naturalisti din Iasi.
- [276] Vrejoiu, Mihnea. (2019). *Rețele neuronale convoluționale, Big Data și Deep Learning în analiza automată de imagini*. Revista Română de Informatică și Automatică. 29. 91-114. 10.33436/v29i1y201909.
- [277] *Neuronul*. Imagine de Clker-Free-Vector_Images, preluată de pe Pixabay. Online: <https://pixabay.com/ro/vectors/neuron-celulei-nervoase-axon-296581/>.
- [278] Dan – Marius Dobrea. *Curs "Tehnici de inteligență computațională. Aplicații în electronică și biomedicină"*, capitolul "Rețele neuronale artificiale". Universitatea Tehnică "Gheorghe Asachi" Iași, anul IV, Facultatea de Electronică, Telecomunicații și Tehnologie Informației. Online: http://www.etc.tuiasi.ro/cin/Courses/Epiom/Course/Curs_Rtizrt123XuAsdr/Capitol%20RNA.pdf
- [279] A. Chowdhury, A. Ayman, S. Dey, M. Sarkar and A. I. Arka (2017). *Simulations of threshold logic unit problems using memristor based synapses and CMOS neuron*. 2017 3rd International Conference on Electrical Information and Communication Technology (EICT), Khulna, 2017, pp. 1-4, doi: 10.1109/EICT.2017.8275126.
- [280] Liu, Qingshan & Wang, Jun. (2011). *A One-Layer Dual Recurrent Neural Network with a Heaviside Step Activation Function for Linear Programming with Its Linear Assignment Application*. 6792. 253-260. 10.1007/978-3-642-21738-8_33.
- [281] Rosenblatt, F. (1958). *The perceptron: a probabilistic model for information storage and organization in the brain*. Psychological review, 65 6, 386-408.
- [282] Goodfellow, Ian; Bengio, Yoshua; Courville, Aaron (2016). *6.5 Back-Propagation and Other Differentiation Algorithms*. Deep Learning. MIT Press. pp. 200–220. ISBN 9780262035613.
- [283] Virgil Tiponut, Cătălin-Daniel Căleanu (2002). *Rețele neuronale. Arhitecturi și algoritmi*. Editura Politehnică, Timișoara. pp. 11-51. ISBN: 973-9389-66-X.
- [284] Cruz, Joseph & Wishart, David. (2007). *Applications of Machine Learning in Cancer Prediction and Prognosis*. Cancer informatics, 2, pp. 59-77. 10.1177/117693510600200030.
- [285] Huang, Gao & Liu, Zhuang & Weinberger, Kilian. (2016). *Densely Connected Convolutional Networks*. 12.
- [286] Y. LeCun et al. (1989). Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation*, vol. 1, no. 4, pp. 541-551, Dec. 1989, doi: 10.1162/neco.1989.1.4.541.
- [287] Fukushima K, Miyake S (1982). Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In: *Competition and cooperation in neural nets*. Springer, pp 267–285.
- [288] Asifullah Khan, Anabia Sohail, Umme Zahoora, Aqsa Saeed Qureshi (2019). *A Survey of the Recent Architectures of Deep Convolutional Neural Networks*. Artificial Intelligence Review, DOI: <https://doi.org/10.1007/s10462-020-09825-6>.
- [289] SlidesGo și Freepik. Online: <https://slidesgo.com/>. Online: <https://arxiv.org/ftp/arxiv/papers/1901/1901.06032.pdf>.
- [290] Ian Goodfellow and Yoshua Bengio and Aaron Courville (2016). *Deep Learning*. MIT Press. p. 326, chapter 9. Online: <https://www.deeplearningbook.org/contents/convnets.html>.
- [291] Afshine Amidi, Shervine Amidi (2018). *Super VIP Cheatsheet: Deep Learning*. Stanford University, Chapter 1, pp: 2-7. Online: <https://stanford.edu/~shervine/teaching/cs-230/cheatsheet-convolutional-neural-networks>.
- [292] Bahri, Mawardi & Ashino, Ryuichi & Vaillancourt, Rémi. (2013). *Convolution Theorems for Quaternion Fourier Transform: Properties and Applications*. Abstract and Applied Analysis. 2013. 10.1155/2013/162769.

- [293] S. Albawi, T. A. Mohammed and S. Al-Zawi (2017). *Understanding of a convolutional neural network*. 2017 International Conference on Engineering and Technology (ICET), Antalya, 2017, pp. 1-6, doi: 10.1109/ICEngTechnol.2017.8308186.
- [294] O'Shea, Keiron & Nash, Ryan. (2015). *An Introduction to Convolutional Neural Networks*. ArXiv e-prints. Online: <https://arxiv.org/pdf/1511.08458.pdf>.
- [295] W. Ausawalaithong, A. Thirach, S. Marukatat and T. Wilaiprasitporn (2018). *Automatic Lung Cancer Prediction from Chest X-ray Images Using the Deep Learning Approach*. 2018 11th Biomedical Engineering International Conference (BMEiCON), Chiang Mai, 2018, pp. 1-5, doi: 10.1109/BMEiCON.2018.8609997.
- [296] Lee, Honglak & Grosse, Roger & Ranganath, Rajesh & Ng, Andrew. (2009). *Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations*. Proceedings of the 26th International Conference on Machine Learning, ICML 2009. 77. 10.1145/1553374.1553453.
- [297] ROVISLaboratory, *Machine Learning, Curs 8 – Convolutional Neural Network*. Online: https://rovislab.com/courses/ml/08_NeuralNetworks-CNN/08_Retele_Neurale_Convolutionale.pdf.
- [298] Rawat, W., & Wang, Z. (2017). *Deep convolutional neural networks for image classification: A comprehensive review*. Neural computation, 29(9), 2352-2449.
- [299] Setul de date claritin_october_twitter_side_effects disponibil pe platforma data.world. Online: <https://data.world/crowdfunder/claritin-twitter/workspace/>.
- [300] Mike Planty, Deven Carlson (2015). *Understanding Education Indicators: A Practical Primer for Research and Policy*. Teachers College Press, Columbia University New York and London, 2015. ISBN: 9780807771723.
- [301] *Quick Guide to Education Indicators for SDG 4*. UNESCO Institute for Statistics. Canada, 2018. Online: <http://uis.unesco.org/sites/default/files/documents/quick-guide-education-indicators-sdg4-2018-en.pdf>.
- [302] Barbu, Dragos. (2019). *Soluții de prelucrare specifice Big Data*. Revista Română de Informatică și Automatică. DOI: 29. 10.33436/v29i2y201903.
- [303] Dan Munteanu, Nicoleta Munteanu (2020). *Sentiment Analysis Based on Deep Learning Techniques Applied to Children in Logical Games from Nonformal Education*. International Scientific Conference eLearning and Software for Education, Volume 1, pp 57-65, Bucharest, April 30 - May 1, 2020, DOI: 10.12753/2066-026X-20-007.
- [304] Gutu, G., Dascalu, M., Trausan-Matu, S., & Dessus, P. (2016). *ReaderBench goes online: A Comprehension-Centered Framework for Educational Purposes*. RoCHI.
- [305] Henning, Erik. (2020). *Testare automată prin Vedere Artificială*. 10.13140/RG.2.2.11501.82404.d
- [306] S. L. Jurj, F. Opritoiu and M. Vladutiu (2018). *Identification of Traditional Motifs Using Convolutional Neural Networks*. 2018 IEEE 24th International Symposium for Design and Technology in Electronic Packaging (SIITME), 2018, pp. 191-196, doi: 10.1109/SIITME.2018.8599199.
- [307] Orhei C, Vert S, Mocofan M, Vasiu R (2021). *End-To-End Computer Vision Framework: An Open-Source Platform for Research and Education*. Sensors. 2021; 21(11):3691. <https://doi.org/10.3390/s21113691>.
- [308] Ionescu, Bogdan & Coquin, Didier & Lambert, Patrick & Vasile, Buzuloiu. (2005). *Dynamic Hand Gesture Recognition Using the Skeleton of the Hand*. EURASIP Journal on Advances in Signal Processing. 2005. 10.1155/ASP.2005.2101.
- [309] Shubhra Kanti Karmaker ("Santu"), Md. Mahadi Hassan, Micah J. Smith, Lei Xu, Chengxiang Zhai, and Kalyan Veeramachaneni. (2021). *AutoML to Date and Beyond: Challenges and Opportunities*. ACM Comput. Surv. 54, 8, Article 175 (November 2022), 36 pages. <https://doi.org/10.1145/3470918>
- [310] AutoML. Online: <https://www.automl.org/automl/>
- [311] Elsken, T., Metzen, J. H., & Hutter, F. (2019). *Neural architecture search: A survey*. The Journal of Machine Learning Research, 20(1), 1997-2017.

- [312] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas (2016). *Taking the human out of the loop: A review of bayesian optimization*. Proceedings of the IEEE, 104(1):148–175, Jan 2016. ISSN 0018-9219. DOI: 10.1109/JPROC.2015.2494218.
- [313] Sutton, R.S. (1992). Introduction: The Challenge of Reinforcement Learning. In: Sutton, R.S. (eds) *Reinforcement Learning. The Springer International Series in Engineering and Computer Science*, vol 173. Springer, Boston, MA. https://doi.org/10.1007/978-1-4615-3618-5_1
- [314] *Migrarea datelor din Amazon redshift folosind BigQuery Data Transfer Service*. Documentație oficială Google. Online: <https://cloud.google.com/bigquery-transfer/docs/redshift-migration>.
- [315] P. Pierleoni, R. Concetti, A. Belli and L. Palma (2020). *Amazon, Google and Microsoft Solutions for IoT: Architectures and a Performance Comparison*. IEEE Access, vol. 8, pp. 5455-5470, 2020, doi: 10.1109/ACCESS.2019.2961511.
- [316] Sfetcu, Nicolae. (2019). *Modele ale emoțiilor*. DOI: 10.13140/RG.2.2.24223.30884.
- [317] Marhan, Ana Maria. (2008). Interacțiunea conversațională în sistemele colaborative pe web. St. Trăușan-Matu (eds.). *Capitolul 4: Modalități de reprezentare a emoțiilor în comunicarea bazată pe text* (pp.81-104). București: MatrixRom, 2008, p. 81-104 ISBN 978-973-755-393-5.
- [318] Ekman, P., Friesen, W. V., & Ellsworth, P. (2013). *Emotion in the human face: Guidelines for research and an integration of findings* (Vol. 11). Elsevier.
- [319] Credit pentru imaginea "Roata emoțiilor": NicolaeSfetcu, CC0, via Wikimedia Commons. Link de descărcare: https://commons.wikimedia.org/wiki/File:Roata_emo%C8%9Bilor_lui_Plutchik.png.
- [320] Documentație oficială Orange. Online: <https://orangedatamining.com/widget-catalog/text-mining/preprocesstext/>.
- [321] *Stemming and Lemmatization*. Stanford.Edu. Online: <https://nlp.stanford.edu/IR-book/html/htmledition/stemming-and-lemmatization-1.html>. Accesat la 3 februarie 2023.
- [322] P. Han, S. Shen, D. Wang and Y. Liu, "The influence of word normalization in English document clustering," 2012 IEEE International Conference on Computer Science and Automation Engineering (CSAE), Zhangjiajie, China, 2012, pp. 116-120, doi: 10.1109/CSAE.2012.6272740.
- [323] Banga, Ritu & Mehndiratta, Pulkit. (2017). *Tagging Efficiency Analysis on Part of Speech Taggers*. 2017 International Conference on Information Technology (ICIT), ACM, New York. 264-267. 10.1109/ICIT.2017.57.
- [324] Riedwyl, H., and Schüpbach, M. (1994). *Parquet diagram to plot contingency tables*. In Softstat '93: Advances in Statistical Software, F. Faulbaum (Ed.). New York: Gustav Fischer, 293-299.
- [325] Meyer D., Zeileis A., Hornik K. (2008). *Visualizing Contingency Tables*. In: *Handbook of Data Visualization*. Springer Handbooks Comp. Statistics. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-33037-0_23
- [326] Lee J, Suh T, Roy D, Baucus M (2019). *Emerging Technology and Business Model Innovation: The Case of Artificial Intelligence*. Journal of Open Innovation: Technology, Market, and Complexity. 2019; 5(3):44. <https://doi.org/10.3390/joitmc5030044>
- [327] Davenport, T., Guha, A., Grewal, D. et al. (2020). *How artificial intelligence will change the future of marketing*. J. of the Acad. Mark. Sci. 48, 24–42 (2020). <https://doi.org/10.1007/s11747-019-00696-0>
- [328] Documentație AutoML Vision. Online: <https://cloud.google.com/vision/automl/docs/create-datasets>.
- [329] Credit pentru imaginile utilizate. Pixabay Germania. Online: <https://pixabay.com/ro/images/search/tristete/>.
- [330] Documentație Clarifai. *Modelele Clarifai*. Online: <https://docs.clarifai.com/api-guide/model/clarifai-models>.
- [331] Documentație oficială Clarifai. Concepts. Online: <https://docs.clarifai.com/api-guide/concepts>.
- [332] Documentație oficială Clarifai. Annotations. Online: <https://docs.clarifai.com/api-guide/annotate/annotations>.
- [333] Harari, Yuval (2019). *21 Lessons for the 21st Century*. London, England. ISBN 9781784708283.

[334] Site oficial Google Cloud. Online: <https://cloud.google.com/products>.