

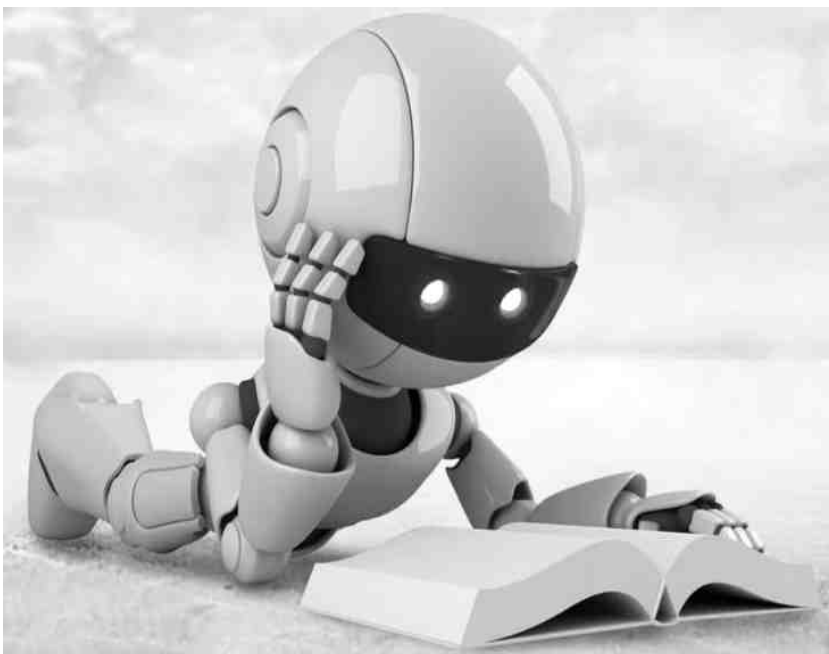


Digitally signed by
Technical Scientific
Library, TUM
Reason: I attest to the
accuracy and integrity of
this document

UNIVERSITATEA TEHNICĂ A MOLDOVEI

ÎNVĂȚARE AUTOMATĂ

SUPORT DE CURS



**Chișinău
2023**

UNIVERSITATEA TEHNICĂ A MOLDOVEI
FACULTATEA CALCULATOARE, INFORMATICĂ ȘI
MICROELECTRONICĂ
DEPARTAMENTUL INFORMATICĂ ȘI INGINERIA
SISTEMELOR

ÎNVĂȚARE AUTOMATĂ

SUPORT DE CURS

Chișinău
Editura „Tehnica-UTM”
2023

CZU 004.85(075.8)

B 60

Lucrarea a fost discutată și aprobată pentru editare la ședința Consiliului Facultății Calculatoare, Informatică și Microelectronică, proces-verbal nr. 6 din 26.04.2023.

Lucrarea de față conține material teoretic însoțit de exemple explicate detaliat referitor la cei mai cunoscuți algoritmi de învățare automată și metodele de aplicare a lor care includ analiza și preprocesarea datelor de învățare, principiile de lucru ale algoritmilor, pașii principali de utilizare a algoritmilor și metodele de evaluare a rezultatelor obținute. În final, este descris un set de instrumente soft WEKA care conține modulele necesare tuturor etapelor de utilizare a algoritmilor de învățare automată și un număr vast de implementări ale algoritmilor dați.

Materialul expus în lucrare este necesar la studierea disciplinei *Învățarea automată* inclusă în planul de învățământ pentru programele de studii superioare de licență *Infomatica aplicată și Managementul aplicațiilor infromaționale*, Facultatea Calculatoare, Informatică și Microelectronică, UTM. De asemenea, materialul va servi drept suport la îndeplinirea lucrărilor de laborator la disciplina menționată.

Autor: conf. univ., dr. Victoria Bobicev

Redactor responsabil: conf. univ., dr. Viorica Sudacevschi

Recenzent: conf. univ., dr. Ilie Coandă

DESCRIEREA CIP A CAMEREI NAȚIONALE A CĂRȚII DIN RM

Bobicev, Victoria.

Învățare automată: Suport de curs / Victoria Bobicev; redactor responsabil: Viorica Sudacevschi; Universitatea Tehnică a Moldovei, Facultatea Calculatoare, Informatică și Microelectronică, Departamentul Informatică și Ingineria Sistemelor.

– Chișinău: Tehnica-UTM, 2023. – 127 p.: fig., tab.

Aut. indicat pe verso f. de tit.

– Referințe bibliogr.: p. 126-127 (12 tit.). – 50 ex.

CUPRINS

Introducere	5
1. Noțiuni de bază în învățarea automată.....	7
1.1. Preprocesarea datelor.....	8
1.2. Tipuri de învățare automată	12
1.3. Învățare supervizată.....	13
1.4. Algoritmi de învățare supervizată.....	15
Subiecte pentru verificarea cunoștințelor:	17
2. Regresia liniară	18
2.1. Noțiuni de bază privind regresia liniară.	18
2.2. Regresia liniară simplă	19
2.3. Metodă de estimare a liniei optime	22
2.4. Intervalul de încredere pentru regresia liniară	24
2.5. Regresia liniară multivariată	27
2.6. Regresia logistică	28
Subiecte pentru verificarea cunoștințelor:	32
3. Metoda Bayes	33
3.1. Formula Bayes.....	33
3.2. Formula Bayes în învățarea automată.....	36
3.3. Formularea formalizată a clasificatorului Bayesian	39
3.4. Bayes naiv.....	41
3.5. Exemplu de demonstrare a metodei Bayes.	42
Subiecte pentru verificarea cunoștințelor:	46
4. Metoda K-NN.....	47
4.1. Componentele sistemului de învățare bazat pe instanțe	48
4.2. Exempu de calculare a distanței:	49

4.3.	Distanța Levenshtein	51
4.4.	Câștig de informație.	58
4.5.	Entropia	59
4.6.	Ponderarea caracteristicilor	64
	Subiecte pentru verificarea cunoștințelor:	70
5.	Metoda arborilor de decizie	71
5.1.	Noțiuni inițiale privind arborele de decizie.....	71
5.2.	Metoda arborilor de decizie în învățarea automată.	73
5.3.	Măsuri de selecție a atributelor.	74
5.4.	Algoritmul ID3	74
	Subiecte pentru verificarea cunoștințelor:	82
6.	Evaluarea rezultatelor	84
6.1.	Validarea încrucișată	85
6.2.	Acuratețea clasificării	87
6.3.	Matricia de confuzie	88
6.4.	Receiver Operating Characteristics	91
	Subiecte pentru verificarea cunoștințelor:	96
7.	Setul de instrumente WEKA	97
7.1.	Descărcare și instalare WEKA	98
7.2.	Datele de intrare WEKA.....	102
7.3.	Fila Preprocess în WEKA.....	105
7.4.	Fila Classify în WEKA.....	108
7.5.	Fila Cluster din WEKA	116
7.6.	Fila Associate din WEKA	119
7.7.	Fila Vizualize în WEKA.....	122
7.8.	Alte instrumente propuse de WEKA	123
	Subiecte pentru verificarea cunoștințelor:	125
	BIBLIOGRAFIE	126

INTRODUCERE

Învățarea automată (Machine Learning - ML) este o ramură a inteligenței artificiale care se ocupă cu dezvoltarea algoritmilor și modelelor ce permit sistemelor de calculatoare să învețe și să-și îmbunătățească performanța în mod automat, fără a fi explicit programate pentru fiecare sarcină.

Metodele de învățare automată au cunoscut o semnificativă popularitate și aplicabilitate în ultimii ani datorită progreselor tehnologice și a volumului mare de date disponibile [1].

Lucrarea conține descrierea metodelor de învățare automată. Din mulțimea metodelor sunt selectate cele mai cunoscute și activ utilizate. Materialul teoretic este însoțit de exemple explicate detaliat referitor la cei mai cunoscuți algoritmi de învățare automată și metodele de aplicare a lor care includ analiza și preprocesarea datelor de învățare, principiile de lucru ale algoritmilor, pașii principali de utilizare a algoritmilor și metodele de evaluare a rezultatelor obținute. În final, este descris un set de instrumente soft WEKA care conține modulele necesare pentru toate etapele de utilizare a algoritmilor de învățare automată și un număr vast de implementări ale algoritmilor dați.

Materialul expus în lucrare este necesar la studierea disciplinei *Învățarea automată* inclusă în planul de învățământ pentru programele de studii superioare de licență *Infomatica aplicată* și *Managementul aplicațiilor infromaționale*, Facultatea Calculatoare, Informatică și Microelectronică, UTM. De asemenea, materialul va servi drept suport la îndeplinirea lucrărilor de laborator la disciplina menționată.

Scopul cursului este însușirea de către studenți a principiilor de învățare automată, tehnicilor de preprocesare a datelor, utilizării algoritmilor în practică, precum și evaluării rezultatelor obținute.

Obiectivele principale ale cursului *Învățarea automată* constau în formarea la studenți a următoarelor abilități:

- cunoașterea și înțelegerea bazelor matematice ale algoritmilor de învățare automată;
- dezvoltarea abilităților de utilizare a algoritmilor de învățare automată;
- aplicarea tehnicilor de preprocesare a datelor de antrenare;
- proiectarea, modelarea și experimentarea, folosind diverși algoritmi de învățare automată;
- sporirea interesului față de cercetarea în domeniul învățării automate.

BIBLIOGRAFIE

1. Mircea-Bogdan Rădac. Tehnici de învățare automată cu aplicații. București: Editura Politehnica, 2019. - 234 p.
2. Florin Leon. Inteligența artificială: raționament probabilistic, tehnici de clasificare Tehnopress. Iași, 2012.
3. Quarteroni R., Sacco F., Saleri. Matematică numerică. Ediția a IV-a. Milano: Springer Verlag, 2014.
4. Trevor Hastie, Robert Tibshirani, Jerome Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition (Springer Series in Statistics) 2nd Edition, Springer, 2016.
5. Stulp Freek and Olivier Sigaud. Many Regression Algorithms, One Unified Model: A Review. Neural Networks, vol. 69, 2015, p. 60-79.
6. Hosmer David. Applied logistic regression. Hoboken. New Jersey: Wiley, 2013.
7. Stone J.V. "Bayes' Rule: A Tutorial Introduction to Bayesian Analysis", Sebtel Press, England, 2013.
8. Peter M. Lee. Bayesian Statistics: An Introduction, Wiley, 4th Edition, 2012. - 496 p.
9. Walter Daelemans, Jakub Zavrel. Ko van der Sloot, 2018, TiMBL: Tilburg Memory-Based Learner, version 6.4, Reference Guide, ILK Technical Report – ILK 11-01, University of Antwerp, sursa electronică,
https://linguistics.byu.edu/classes/ling581dl/Timbl_6.4_Manual.pdf, accesat 05.07.2023.
10. Ian H. Witten, Eibe Frank, Mark A. Hall, Cristopher J. Pal. Data Mining: Practical Machine Learning Tools and Techniques, 4th Edition. Morgan Kaufmann, 2016.

11. Eric Sven Ristad, Peter N. Yianilos. Learning String Edit Distance. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, Volume: 20, Issue: 5, Publisher: IEEE.

12. Eibe Frank, Mark A. Hall, Ian H. Witten, 2013. The WEKA Workbench, online Appendix for „Data Mining: Practical Machine Learning Tools and Techniques, 4th Edition” 4th Edition. Morgan Kaufmann., sirsas elektronika, https://www.cs.waikato.ac.nz/ml/weka/Witten_et_al_2016_appendix.pdf, accessed 5.07.2023.