

FIȘA DISCIPLINEI/MODULULUI

 MD-2068, CHIȘINĂU, STR. STUDENȚILOR, 9/7, TEL: 022 50-99-63, www.utm.md
Machine Learning și Data Mining
1. Date despre disciplină/modul

Facultatea	CIM				
Departamentul	Ingineria Software și Automatică				
Ciclul de studii	Masterat				
Programul de studii	Știința Datelor				
Anul de studii	Semestrul	Tip de evaluare	Categoria formativă	Categoria de opționalitate	Credite ECTS
Anul I (<i>învățământ cu fr.</i>)	1	E	F	O	5

2. Timpul total estimat

Total ore în planul de învățământ	Din care				
	Ore auditoriale		Lucrul individual		
	Curs	Laborator/ seminar	Proiect de an	Studiul materialului teoretic	Pregătire aplicații
Învățământ cu frecvență	20	20	-	110	-

3. Precondiții de acces la disciplină/modul

Conform planului de învățământ	Algebra liniară, Matematica superioară, Teoria probabilității și a informației, Statistica, Modele matematice și optimizări, Analiza exploratorie a datelor și modelări, Vizualizarea datelor
Conform competențelor	Abilități în manipularea, preprocesarea și vizualizarea datelor. Gândire critică și analitică cu abilități de rezolvare a problemelor. Deprinderi de autoînvățare. Înțelegerea contextului de afaceri, a modului în care ML se integrează în afaceri și în aplicațiile din lumea reală.

4. Condiții de desfășurare a procesului educațional pentru

Curs	Pentru prezentarea materialului teoretic în sala de curs este nevoie de proiector și calculator. Nu vor fi tolerate întârzierile studenților, precum și convorbirile telefonice în timpul cursului.
Laborator/ seminar	Studenții vor efectua lucrări de laborator sub îndrumarea profesorului și a asistentului și vor perfecta rapoartele în conformitate cu instrucțiunile metodice. Termenul de susținere a lucrărilor de laborator este de o săptămână de la finalizarea acestora. În cazul în care lucrarea este depusă cu întârziere, se va aplica o penalizare de 1 punct per săptămână de întârziere.

5. Competențe specifice acumulate

Competențe profesionale	CPM1 Elaborarea și proiectarea arhitecturii sistemului CPM2 Monitorizarea tendințelor tehnologice. Inovație. Dezvoltarea durabilă. CPM3 Dezvoltarea aplicațiilor. Integrarea componentelor. Ingineria sistemelor. CPM 5. Îmbunătățirea proceselor.
Competențe transversale	CTM1 Autonomie și responsabilitate CTM2 Interacțiune socială CTM3 Dezvoltare profesională și personală

6. Obiectivele disciplinei/modulului

Obiectiv general	Formarea unei baze solide în domeniul învățării automate și al mineritului de date, permițându-le studenților să aplice aceste tehnici în mod eficient în diverse sfere de activitate și să ia decizii informate, bazate pe date.
Obiectivele specifice	În vederea atingerii obiectivului general este necesar de dezvoltat cunoștințe și abilități în: <ul style="list-style-type: none"> • Învățarea automată (ML): înțelegerea conceptelor și principiilor fundamentale ale învățării automate. • Prelucrarea prealabilă a datelor relevante: transformarea datelor, scalarea și

	<p>normalizarea caracteristicilor.</p> <ul style="list-style-type: none"> • Algoritmi de învățare supravegheată: algoritmi de învățare supravegheată, cum ar fi regresia, arborii de decizie, mașinile de suport vectorial etc. • Evaluarea și selectarea algoritmilor adecvați pentru probleme specifice. • Algoritmi de învățare nesupravegheată: tehnici de învățare nesupravegheată - gruparea, reducerea dimensiunilor și extragerea regulilor de asociere. • Metrici de evaluare și selectare a modelelor: Înțelegerea și aplicarea parametrilor de evaluare adecvați pentru diferite tipuri de sarcini de învățare automată (de exemplu, acuratețe, precizie, rechemare, scor F1). • Tehnici de selecție și validare a modelelor (validarea încrucișată, divizarea antrenare-testare). • Conceptele și tehnicile fundamentale ale mineritului de date, inclusiv gruparea, extragerea regulilor de asociere și detectarea valorilor aberante. • Aplicații din lumea reală: tehnicile de învățare automată și de minerit de date la probleme și seturi de date din lumea reală. • Formularea problemelor din categoria preprocesarea datelor și alegerea algoritmilor adecvați pentru rezolvarea lor. • ML etic și responsabil: considerațiile etice privind învățarea automată și extragerea datelor, inclusiv prejudecata, corectitudinea și confidențialitatea. • Dezvoltarea și implementarea responsabile a inteligenței artificiale. • Programare și instrumente practice: Dezvoltarea abilităților practice de programare în Python și dobândirea de competențe în utilizarea bibliotecilor și a platformelor relevante. • Experiența în utilizarea instrumentelor de minerit de date și ML, cum ar fi scikit-learn, TensorFlow și PyTorch.
--	---

7. Conținutul disciplinei/modulului

Tematica activităților didactice	Numărul de ore
	învățământ cu frecvență
Tematica cursurilor	
T1. Introducere în Învățarea supravegheată. Seturi de date de instruire, testare și validare. Compromisul dintre bias și varianță. Alegerea unui model: Criteriul informațional bayesian (BIC) și criteriul informațional Akaike (AIC).	2
T2. Modele de predicție. Modele de regresie multivariate: de la cauzalitate la predicții. Alegerea predictorilor. Metode de construire a modelelor: All-in, Stepwise Regression, Compararea scorurilor. Funcția de pierdere OLS și problema minimizării.	2
T3. Modele de predicție. Suport Vector Regression (SVR). Intuiție. Tub insensibil la erori. Variabile slabe. Funcția de pierdere SVR.	2
T4. Modele de predicție. Regresia arborelui de decizie și regresia pădurilor aleatorii. Intuiție. Calcularea predicțiilor unui arbore de decizie pentru un exemplu aplicativ. Păduri aleatorii și învățare în ansamblu. Validarea încrucișată K-Fold.	2
T5. Modele de clasificare. Regresie logistică și arbore de decizie pentru clasificări. Recapitulare a regresiei logistice. Pro și contra regresiei logistice. Arbori de decizie și păduri Radom pentru clasificare. Entropie, indicele Gini, câștig de informație. Tipuri de arbori. Noduri, frunze și adâncime.	2
T6. Modele de clasificare. Algoritmul K Nearest Neighbors (K-NN). Intuiție. Metrici de distanță (de exemplu, euclidiană). Avantaje și dezavantaje ale K-NN. K-NN ponderat.	2
T7. Modele de clasificare. Algoritmul Naive Bayes (NB). Teorema Bayes. Exemplu de calcul. Probabilități posterioare și prioritară, verosimilitate și verosimilitate marginală. Calculul predicțiilor Naive Bayes. Pro și contra NB.	2
T8. Introducere în învățarea nesupravegheată. Învățarea supravegheată versus învățarea nesupravegheată. Clusterizarea. K-means. Metoda cotului. K-Means	2

Tematica activităților didactice	Numărul de ore
	învățământ cu frecvență
++. Clusterizarea ierarhică. Dendograme. Exemplu de grupare ierarhică folosind dendograme.	
T9. Învățarea regulilor de asociere. Algoritmii Apriori și Eclat. Noțiunile de suport, încredere și ridicare în Apriori și Eclat.	2
T10. Învățarea forțată. Algoritmul Upper Confidence Bound. Problema banditului cu mai multe brațe. Recompensa medie și intervalul de încredere. Exemplu. Algoritmul de eșantionare Thompson. Recompensa și distribuția Bemouli. Inferența bayesiană în algoritmul de eșantionare Thompson. Exemplu.	2
Total curs:	20
Tematica lucrărilor de laborator	
Lucrarea de laborator nr. 1. Implementarea regresiei liniare multivariate în Python pentru a face predicții.	2
Lucrarea de laborator nr. 2. SVR în Python.	2
Lucrarea de laborator nr. 3. Arbori de decizie și păduri aleatorii în Python.	2
Lucrarea de laborator nr. 4. K-NN și Naive Bayes în Python	2
Lucrarea de laborator nr. 5. Compararea modelelor și criteriile de selecție.	2
Lucrarea de laborator nr. 6. Clusterizare. K-means și gruparea ierarhică în Python.	4
Lucrarea de laborator nr. 7. Algoritmi Apriori și Eclat în Python.	2
Lucrarea de laborator nr. 8. Thompson Sampling și algoritmi UCB în Python.	4
Total lucrări de laborator:	20

8. Referințe bibliografice

Principale	<ol style="list-style-type: none"> 1. Ian Goodfellow, Yoshua Bengio and Aaron Courville, Deep Learning, 2016, MIT Press, http://www.deeplearningbook.org (Chapters 1-5) 2. O. Theobald, 2017, Machine Learning For Absolute Beginners: A Plain English Introduction, Second Edition (AI, Data Science, Python & Statistics for Beginners), Scatterplot Press 3. Educational materials and bibliographic sources on fcim's else platform: https://else.fcim.utm.md/course/view.php?id=702
Suplimentare	<ol style="list-style-type: none"> 4. Andriy Burkov, 2019, The Hundred-Page Machine Learning Book, ISBN-10: 199957950X, ISBN-13 : 978-1999579500 5. John D. Kelleher and Brian Mac Namee, 2015, Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies, MIT-Press, ISBN-10 : 0262029448, ISBN-13 978-0262029445. 6. Aurélien Géron, 2019, Hands-on machine learning with scikit-learn, keras, and tensorflow, 2nd edition, o'reilly media, inc. isbn: 9781492032649; https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632/

9. Evaluare

Periodică		Curentă	Studiu individual	Proiect/teză	Examen
EP 1	EP 2				
Învățământ cu frecvență					
15%	15%	15%	15%		40%
Standard minim de performanță Prezența la lecții; activitatea și calitatea pregătirii la / pentru prelegeri și lucrări de laborator; Obținerea notei minime de „5” la fiecare dintre atestări și lucrări de laborator; Demonstrarea în lucrarea de examinare finală a cunoașterii conținutului teoretic și abilităților practice.					

10. Criterii de evaluare

Activitate	Componente evaluare	Metodă de evaluare, Criterii de evaluare	Pondere în nota finală a activității	Ponderea în evaluarea disciplinei

Activitate	Componente evaluare	Metodă de evaluare, Criterii de evaluare	Pondere în nota finală a activității	Ponderea în evaluarea disciplinei
Învățământ cu frecvență				
Evaluare periodică I	Conținut teoretic, teme 1 - 4	Test	100%	15%
Evaluare periodică II	Conținut teoretic, teme 5 - 8	Test	100%	15%
Evaluare curentă	Activitatea practică	Discuții în cadrul laboratoarelor	50%	15%
		Rapoarte pentru fiecare lucrare de laborator	50%	
Studiul individual	Cercetare la temă	Prezentare/discurs public	100%	15%
Evaluarea finală	Conținut teoretic și practic	Written Examination. Grading according to grading scale.	100%	40%