

MACHINE LEARNING ȘI DATA MINING

1. Date despre disciplină/modul

Facultatea	Calculatoare, Informatică și Microelectronică				
Departamentul	Inginerie Software și Automatică				
Ciclul de studii	Ciclul II, Studii superioare de master				
Programul de studii	Ingineria software				
Anul de studii	Semestrul	Tip de evaluare	Categoria formativă	Categoria de opționalitate	Credite ECTS
Anul I (învățământ cu frecvență)	I	E	F – unite de curs fundamentală	O - unitate de curs obligatorie	5

2. Timpul total estimat

Total ore în planul de învățământ		Din care				
		Ore auditoriale		Lucrul individual		
		Curs	Practice	Proiect de an	Studiul materialului teoretic	Pregătire aplicații
Învățământ cu frecvență	150	20	20	-	60	50

3. Precondiții de acces la disciplină/modul

Conform planului de învățământ	Finalizarea cursurilor: Bazele programării, Structuri de date și algoritmi, Matematică aplicată, Statistici. Promovarea modulelor relevante pentru inteligența artificială sau analiza datelor.
Conform competențelor	- Cunoștințe intermediare de programare în Python (Pandas, NumPy, Matplotlib). - Bazele probabilităților și statisticii (probabilități condiționate, regresie liniară). - Abilități de procesare și analiză a datelor (curățare, transformare, interpretare). - Experiență în implementarea algoritmilor și optimizarea soluțiilor.

4. Condiții de desfășurare a procesului educațional pentru

Curs	- Sală dotată cu proiector, calculator, internet și tablă interactivă. - Acces la materiale didactice și platforme de învățare
Practice	- Laborator IT cu calculatoare echipate cu Python și biblioteci relevante. - Seturi de date pentru exerciții practice și acces la platforme cloud. - Suport tehnic și didactic pentru implementarea algoritmilor.

5. Competențe specifice acumulate

Competențe profesionale	CPM 1. Elaborarea și proiectarea arhitecturii K1 Modele de arhitectura, metodologii și instrumente de proiectare a sistemelor K4 Arhitectura întreprinderii și standardele interne ale companiei K5 Noi tehnologii emergente (de exemplu, sisteme distribuite, modele de virtualizare, seturi de date, sisteme mobile) S2 Utilizează cunoștințele sale tehnologice din diferite domenii pentru a elabora și implementa arhitectura întreprinderii S3 Înțelege obiectivele companiei care au impact asupra componentelor arhitecturii (date, aplicații, securitate, dezvoltare etc.) S5 Dezvoltă modele de design și modele arhitecturale pentru a ajuta analiștii sistemului în proiectarea unor aplicații coerente CPM 2. Monitorizarea tendințelor tehnologice. Inovație. Dezvoltarea durabilă K1 Tehnologiile existente și emergente și aplicațiile lor relevante pe piață K2 Obiectivele, tendințele și nevoile business-ului, societății și cercetării K3 Sursele relevante de informații (de exemplu, reviste, conferințe și evenimente, lideri de opinie, forumuri on-line etc.) K4 Abordările concrete ale programelor de cercetare aplicate K5 Tehnicile procesului de inovare S1 Monitorizează sursele de informații și urmărește continuu pe cele mai promițătoare S2 Identifică vânzătorii și furnizorii de cele mai promițătoare soluții; evaluează, justifică și le
--------------------------------	--

	propune pe cele mai potrivite S3 Identifică avantajele și îmbunătățirile care le aduce adoptarea tehnologiilor emergente S6 Stăpânește principalele constrângeri ale standardelor internaționale referitor la sustenabilitatea TIC CPM 3. Dezvoltarea aplicațiilor. Integrarea componentelor. Ingineria sistemelor K1 Programe / module adecvate, SGBD și limbaje de programare adecvate. Tehnologii de ultimă oră K3 Impactul integrării unui sistem asupra organizației sau a sistemului existent K4 Tehnici de interfațare între module, sisteme și componente K7 Componente hardware, instrumente și arhitecturi hardware K8 Proiectarea funcțională și tehnică S1 Aplică arhitecturi software și / sau hardware adecvate S2 Măsoară performanța sistemului înainte, în timpul și după integrarea sistemului S3 Identifică și înregistrează activitățile, problemele și măsurile corective legate de întreținere S4 Adaptează nevoile clienților la produsele existente S5 Securizează și face backup-ul datelor pentru a asigura integritatea lor în timpul integrării datelor sau a sistemului S9 Aplică modele de date, procese, pentru a se dezvolta eficient și productiv CPM 5. Îmbunătățirea proceselor K1. Metode de cercetare, comparare și metode de măsurare K2 Metode de evaluare, proiectare și implementare K4 Dezvoltările/Evoluțiile relevante în domeniul TIC (de exemplu, virtualizarea, datele deschise etc.) și impactul lor potențial asupra proceselor K5 Specificitatea tehnologiilor web, cloud și mobile S1 Redactează, documentează și cataloghează procesele și procedurile esențiale S2 Propune modificări ale procesului pentru a facilita și raționaliza îmbunătățirile
Competențe transversale	CTM1. Autonomie și responsabilitate CTM2. Interacțiune socială CTM3. Dezvoltare personală și profesională

6. Obiectivele disciplinei/modulului

Obiectivul general	Dezvoltarea competențelor necesare pentru utilizarea conceptelor, tehnicilor și algoritmilor fundamentali din Machine Learning și Data Mining, cu scopul de a analiza, modela și rezolva probleme reale din diverse domenii utilizând limbaje de programare precum Python.
Obiectivele specifice	• Înțelegerea și aplicarea principiilor de învățare supravegheată, incluzând regresia multivariată, regresia logistică, arborii de decizie și alte metode predictive. • Implementarea algoritmilor de învățare supravegheată (cum ar fi K-NN, SVR, Naive Bayes) și evaluarea performanțelor acestora prin utilizarea seturilor de date de instruire, validare și testare. • Aplicarea tehnicilor de învățare nesupravegheată, inclusiv clusterizarea și analiza regulilor de asociere, în vederea identificării modelelor și structurilor ascunse în seturi de date. • Familiarizarea cu algoritmii de învățare prin întărire, cum ar fi Upper Confidence Bound și Thompson Sampling, și utilizarea acestora în rezolvarea problemelor aplicative. • Utilizarea limbajului Python pentru a dezvolta soluții practice, incluzând implementarea, validarea și compararea modelelor de Machine Learning.

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.	2

Tematica activităților didactice	Numărul de ore
	învățământ cu frecvență
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.	
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 ++. Clusterizarea ierarhică. Dendograme. Exemplu de grupare ierarhică folosind dendograme.	2
T9. Învățarea regulilor de asociere. Algoritmii Apriori și Eclat. Noțiunile de suport, încredere și ridicare în Apriori și Eclat.	2
T10. Învățarea prin întărire. Algoritmul Upper Confidence Bound. Problema banditului cu mai multe brațe. Recompensa medie și intervalul de încredere. Exemplu. Algoritmul de eșantionare Thompson. Recompensa și distribuția Bernouli. Inferența bayesiană în algoritmul de eșantionare Thompson. Exemplu.	2
Total curs:	20
Tematica lucrărilor practice	
LP1. Implementarea regresiei liniare multivariată în Python pentru a face predicții.	2
LP2. SVR în Python.	2
LP3. Arbori de decizie și păduri aleatorii în Python.	2
LP4. K-NN și Naive Bayes în Python.	2
LP5. Compararea modelelor și criterii de selecție.	4
LP6. Clusterizare. K-means și gruparea ierarhică în Python.	2
LP7. Algoritmi Apriori și Eclat în Python.	4
LP8. Thompson Sampling și algoritmii UCB în Python.	2
Total practice:	20

8. Referințe bibliografice

Principale	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 4. Andriy Burkov, 2019, The Hundred-Page Machine Learning Book, ISBN-10: 199957950X, ISBN-13 : 978-1999579500
Suplimentare	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. Utilizarea IA generativă

Permisiunea de utilizare	Utilizarea IA generative în cadrul temelor și proiectelor este permisă, cu condiția ca studenții să respecte următoarele reguli: • IA generativă poate fi utilizată pentru generarea de idei, structuri de text sau cod, dar toate materialele generate trebuie să fie revizuite și ajustate de către student pentru a se asigura că acestea corespund cerințelor academice. • Orice utilizare a IA generative trebuie să fie declarată în secțiunea de apendice a fiecărei lucrări, folosind fraza: "În timpul pregătirii acestei lucrări, autorul a utilizat [NUME INSTRUMENT / SERVICIU] în scopul [MOTIV]. După utilizarea acestui instrument/serviciu, autorul a revizuit și editat conținutul după cum a fost necesar și își asumă întreaga responsabilitate pentru conținutul lucrării."
---------------------------------	---

Restricții de utilizare	Studentii nu trebuie să considere IA generativă ca o sursă de încredere pentru informații, deoarece nu oferă referințe clare sau surse documentate. - Nu este permisă citarea directă a conținutului generat de IA în lucrările academice ca și cum ar fi sursă primară. - Activitățile în care este interzis utilizarea IA generativă sunt specificare de profesor și sunt de regulă evaluări intermediare și finale sau care nu presupun activități de dezvoltare a competențelor profesionale.
--------------------------------	---

10. 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 practică;					
Obținerea notei minime de „5” la fiecare dintre atestări și lucrări practice;					
Demonstrarea în lucrarea de examinare finală a cunoașterii conținutului teoretic al cursului și limbajului Python/R.					

11. Criterii de evaluare

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