

FIȘA DISCIPLINEI/MODULULUI

 MD-2068, CHIȘINĂU, STR. STUDENȚILOR, 9/7, TEL: 022 50-99-63, www.utm.md
Analiza exploratorie și modelarea datelor
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 frecvență</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	Matematici (Algebra liniară, Calculul diferențial și integral, Teoria Probabilităților, Statistica matematică), Structuri de date și algoritmi, Programări (Python sau R), Baze de date, Tehnologii Web.
Conform competențelor	Abilități de lucru și cunoștințe de bază din domeniul datelor, noțiuni statistice de bază, concepte principale din domeniul matematicii superioare (inclusiv din algebra liniară, calculul diferențial și integral, etc), abilități de manipulare a datelor, înțelegerea reprezentării grafice, gândire critică și raționament logic, abilități de programare de bază (concepte de programare în orice limbaj, de preferință Python sau R), concepte de colectare a datelor, cunoștințe privind modul în care sunt stocate și extrase datele din bazele de date, interes și curiozitate pentru analiza datelor, înțelegerea contextului de afaceri, deprinderi de autoinstruire.

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 seminare sub îndrumarea profesorului și a asistentului și vor perfecta rapoartele în conformitate cu instrucțiunile metodice. Termenul de susținere a lucrărilor individuale 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.
Competențe transversale	CT1. Autonomie și responsabilitate CT2. Interacțiune socială CT3. Dezvoltare personală și profesională

6. Obiectivele disciplinei/modulului

Obiectiv general	Dezvoltarea cunoștințelor și formarea abilităților necesare pentru utilizarea eficientă a metodelor și tehnicilor de analiză a datelor în scopul detectării și captării esenței informaționale ascunse în date și al obținerii unei reprezentări simplificate, clare și ușor de interpretat a acestei esențe.
Obiectivele specifice	<p>În vederea atingerii obiectivului general este necesar de dezvoltat cunoștințe și abilități în:</p> <ul style="list-style-type: none"> • Profilarea și rezumarea datelor: tehnici de profilare și rezumare eficientă a seturilor de date. • Vizualizarea datelor: abilități de reprezentare vizuală a datelor utilizând diverse grafice, diagrame, etc. • Identificarea modelelor și tendințelor: identificare tipare, tendințe și relații în cadrul datelor pentru a genera perspective. • Gestionarea datelor lipsă și a datelor aberante: metode de tratare a datelor lipsă și de detectare a valorilor aberante în timpul analizei exploratorii. • Tehnici de analiză statistică: tehnici statistice pentru a analiza datele și a trage concluzii. • Prelucrarea prealabilă a datelor: cunoștințe privind tehnicile de preprocesare a datelor pentru a pregăti datele pentru analiză. • EDA specifică domeniului: tehnici de EDA specifice diferitelor domenii, cum ar fi finanțele, asistența medicală sau marketingul. • Comunicarea constatărilor: abilități de a comunica în mod eficient informațiile și constatările obținute în urma EDA, atât către publicul tehnic și non-tehnic. • Practici de colaborare în domeniul EDA cu colegii în realizarea AEA și împărtășirea informațiilor pentru o înțelegere colectivă. • EDA etică și responsabilă: considerații și responsabilități etice asociate cu efectuarea EDA, inclusiv confidențialitatea și protecția datelor.

7. Conținutul disciplinei/modulului

Tematica activităților didactice	Numărul de ore
Tematica cursurilor	
T1 EDA Subiectul cercetării, motivația și metodele. Statistici descriptive și tehnici grafice pentru explorarea datelor. Definiții ale noțiunilor de bază precum populație, eșantion, observație. Măsuri de tendință centrală (medie, mod, mediană) și proprietățile acestora. Măsuri de variabilitate și dispersie (interval, varianță, abatere standard, asimetrie, kurtosis) și proprietățile acestora. Teorema lui Chebyshev și regula empirică de repartitie pentru distribuțiile normale.	2
T2 Tipuri de date și metode grafice de explorare. Date calitative: Date nominale vs. date ordinale. Date cantitative: Date discrete vs. date continue. Explorarea datelor calitative - Diagrame de frecvență și diagrame circulare. Explorarea datelor cantitative - Histograme, quartile, percentile și măsuri de poziție relativă. Interquartile Range și Box Plots. Date din serii temporale - diagrame liniare și tendințe.	2
T3 Relații între două variabile. Direcția, tipul și relevanța relației. Corelația, covarianța și coeficientul de corelație Pearson. Interpretare și exemple. Corelația falsă.	2
T4. Metoda celor mai mici pătrate. Regresia liniară simplă. Suma pătratelor reziduale și eroarea stocastică. Problema de minimizare a reziduurilor. Exemple.	2
T5. Regresia liniară multivariată și calitatea ajustării. R pătrat, suma totală a pătratelor, suma reziduală a pătratelor și suma explicată a pătratelor.	2
T6. Testarea ipotezelor. Ipotezele nulă și alternativă. Erori de tip I și de tip II. Regula de decizie. Zone de acceptare și de respingere. Statistica T, t-critic și valoarea P a terenului de semnificație. Exemple.	2
T7. Testarea ipotezelor multiple și semnificația regresiei. Testul F, F-critic, grade de libertate.	2

Tematica activităților didactice		Numărul de ore
T8. Specificarea modelului și alegerea predictorilor. Ipoteze OLS. Influența variabilelor omise. Suprimarea interceptului. Regresii polinomiale. Variabile fictive.		2
T9. Regresie logistică. Înțelegerea regresiei logistice și a funcției Logit. Interpretarea rezultatelor regresiei logistice.		2
T10 Serii temporale. Tendință, ciclicitate, sezonabilitate. AR, MA, ARIMA, SARIMA, VARIMA. Diferențiere și staționaritate. Testul Dickey-Fuller (DF) și testul Dickey-Fuller augmentat (ADF). Funcția de autocorelație (ACF) și funcția de autocorelație parțială (PACF).		2
Total curs, ore		20
Tematica lucrărilor practice		
Lucrarea practică nr. 1. Introducere în utilizarea Python în Google Colaboratory - Încărcarea seturilor de date, calculul și explorarea statisticilor descriptive.		2
Lucrarea practică nr. 2. Vizualizarea și analiza datelor utilizând metode grafice în Python.		2
Lucrarea practică nr. 3. Corelații între variabile în Python. Heatmaps și interpretarea lor.		2
Lucrarea practică nr. 4. Implementarea și interpretarea unui model simplu de regresie liniară în Python		2
Lucrarea practică nr. 5. Regresia multivariată în Python și evaluarea bonității de potrivire a modelului.		2
Lucrarea practică nr. 6. Testarea ipotezelor în Python, Interpretarea semnificației statistice.		2
Lucrarea practică nr. 7. Regresia multivariată în Python și testarea comună a semnificației cu ajutorul testului F.		2
Lucrarea practică nr. 8. Testarea specificației modelului și simularea variabilelor omise.		2
Lucrarea practică nr. 9. Regresie logistică în Python. Codificarea variabilelor binare.		2
Lucrarea practică nr. 10. Modelarea seriilor temporale în Python: AR, MA, ARIMA. ACF și PACF. Testul DF și ADF în Python - implementare și interpretare.		2
Total lucrări practice, ore		20

8. Referințe bibliografice

Principale	<ol style="list-style-type: none"> 1. Keller Gerald, Statistics for Management and Economics, South-Western College Publishing; 10th edition, 2014 2. Studenmund Arnold H., Using Econometrics: A practical Guide, Pearson Education, 2014 3. Educational Materials and Bibliographic Sources on FCIM's ELSE Platform: https://else.fcim.utm.md/enrol/index.php?id=701
Suplimentare	<ol style="list-style-type: none"> 1. Wes McKinney, Python for Data Analysis, O'Reilly Media, 3rd Edition, 2022 ISBN: 9781098104009 2. Jake Vanderplas, Python Data Science Handbook: Essential Tools for Working with Data. O'Reilly Media, 2nd Edition, 2023, ISBN-10: 1098121228, ISBN-13: 978-1098121228

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 practice; 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.

10. Criterii de evaluare

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 pe MOODLE	100%	15%
Evaluare periodică II	Conținut teoretic, teme 6-8	Test pe MOODLE	100%	15%
Evaluare curentă	Activitatea practică	Discuții în cadrul seminarelor	50%	15%
		Dosar completat cu Rapoarte pentru fiecare Studiu de caz în discuție	50%	
Studiul individual	Cercetare la temă	Prezentare/discurs public	100%	15%
Evaluarea finală	Conținut teoretic și practic	Examen oral. Notare conform baremului	100%	40%