

ANALIZA EXPLORATORIE A DATELOR

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 (<i>învățământ cu frecvență</i>)	I	E	S – unite de curs de specialitate	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	Pentru a atinge obiectivele disciplinei studenții trebuie să posede cunoștințe din domeniile: 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	- Capacitatea de a aplica metode statistice și tehnici de modelare matematică pentru analiza datelor. - Utilizarea instrumentelor de programare (Python sau R) pentru procesarea și analiza datelor. - Cunoștințe despre structura și organizarea bazelor de date și lucrul cu seturi de date mari. - Abilități în utilizarea reprezentărilor grafice și vizualizării datelor pentru interpretare.

4. Condiții de desfășurare a procesului educațional pentru

Curs	- Procesul educațional se desfășoară în săli dotate cu echipamente de prezentare moderne (videoproector, ecran, acces la internet). - Acces la platforme educaționale pentru materiale de curs, exemple de cod și seturi de date pentru demonstrații. - Utilizarea aplicațiilor statistice și a mediilor interactive de programare.
Practice	- Activitățile practice au loc în laboratoare informatice echipate cu calculatoare performante, conexiune la internet și acces la software specializat pentru analiză și procesare de date. - Realizarea sarcinilor practice folosind seturi de date reale, rezolvarea problemelor statistice și testarea modelelor teoretice în mediul de programare. - Posibilitatea de a lucra în echipă pentru analizarea problemelor complexe și dezvoltarea abilităților de colaborare.

5. Competențe specifice acumulate

Competențe profesionale	CPM 1. Elaborarea și proiectarea arhitecturii K1 Modele de arhitectura, metodologii și instrumente de proiectare a sistemelor K2 Cerințele arhitecturii sistemelor: performanță, mentenabilitate, extensibilitate, scalabilitate, disponibilitate, securitate și accesibilitate K3 Costurile, beneficiile și riscurile unei arhitecturi a sistemului K4 Arhitectura întreprinderii și standardele interne ale companiei K5 Noi tehnologii emergente (de exemplu, sisteme distribuite, modele de virtualizare, seturi de date, sisteme mobile) S1 Oferă expertiză pentru a ajuta la rezolvarea problemelor tehnice complexe și pentru a asigura implementarea celor mai bune soluții de arhitectură 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.) S4 Ajută la comunicarea arhitecturii întreprinderii și a standardelor, principiilor și obiectivelor către diferite echipe implicate 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 K6 Criteriile și indicatorii dezvoltării durabile K7 Responsabilitatea societală corporativă (CSR) a părților interesate din cadrul infrastructurii sistemului informatic 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 S4 Gândește fără idei preconcepute S5 Aplică recomandări în cadrul proiectelor care să sprijine cele mai recente strategii de dezvoltare durabilă 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 K5 Tehnici de testare a integrării K6 Bunele practici de design K7 Componente hardware, instrumente și arhitecturi hardware K8 Proiectarea funcțională și tehnică K9 Bazele securității informației K10 Prototipaje 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 S6 Explică și comunică clientului privind proiectarea / dezvoltarea S7 Lansează și evaluează rezultatele testelor în funcție de specificațiile produsului S8 Proiectează și dezvoltă arhitectura hardware, interfețele utilizatorilor, componentele business software și componentele software integrate S9 Aplică modele de date, procese, pentru a se dezvolta eficient și productiv
Competențe transversale	CTM1. Autonomie și responsabilitate CTM2. Interacțiune socială CTM3. Dezvoltare personală și profesională

6. Obiectivele disciplinei/modulului

Obiectivul 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	Obiective specifice: • înțelegerea conceptelor, metodelor și tehnicilor specifice analizei exploratorii și modelării datelor, precum și a posibilităților de utilizare a acestora în diferite domenii, vizualizării computaționale, generalizării, clasificării, asocierii, clusterizării și învățării statistice; • formarea de abilități privind analiza, modelarea și interpretarea datelor și rezultatelor metodelor și tehnicilor de analiză a datelor utilizate în diferite domenii de activitate; • asigurarea cunoștințelor necesare pentru utilizarea eficientă a metodelor și tehnicilor probabilistice și

	statistico-matematice în scopul fundamentării deciziilor pentru diferite domenii de activitate; • abordarea interdisciplinară a măsurării, cuantificării, analizei și predicției pentru diferite domenii de activitate prin utilizarea unor concepte, metode, instrumente și proceduri specifice teoriei și practicii ingineresti, modelării statistico-matematice și informaticii; • formarea modului de gândire cantitativă a specialistului în știința datelor, dezvoltarea și aprofundarea cunoștințelor acestuia în domeniul măsurării, cuantificării, analizei, evaluării, interpretării și predicției pentru domenii tematice concrete; • formarea de abilități privind înțelegerea și utilizarea instrumentelor software dedicate analizei cantitative și tehnologiilor de analiză și modelare a datelor; • asigurarea de cunoștințe, aptitudini și competențe în domeniul creării, configurării, modificării, testării, executării și mentenanței aplicațiilor de analiză și modelare a datelor; • formarea deprinderilor de lucru în echipă cu respectarea exigențelor etice, deontologice și de colaborare în implementarea proiectelor de grup.
--	--

7. Conținutul disciplinei/modulului

Tematica activităților didactice	Numărul de ore
	învățământ cu frecvență
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, cuartile, 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 intensitatea 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. Exemplu calculării.	2
T5. Regresia liniară multivariată și bonitatea 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. Exemplu.	2
T7. Testarea ipotezelor multiple și semnificația regresiei. Testul F, F-critic, grade de libertate.	2
T8. Specificarea modelului și alegerea predictorilor. Ipoteze OLS (5 OLS Assumptions). 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 de timp. Tendință, ciclicitate, sezonalitate. 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:	20
Tematica lucrărilor practice	
LP1. Introducere în utilizarea Python în Google Colaboratory - Încărcarea seturilor de date, calculul și explorarea statisticilor descriptive.	2
LP2. Vizualizarea și analiza datelor utilizând metode grafice în Python.	2
LP3. Corelații între variabile în Python. Heatmaps și interpretarea lor.	2
LP4. Implementarea și interpretarea unui model de regresie liniară simplă în Python	2
LP5. Regresia multivariată în Python și evaluarea modelului (model fit, F-test & R ²).	2
LP6. Testarea ipotezelor în Python, Interpretarea semnificației statistice.	2
LP7. Regresia multivariată în Python și testarea comună a semnificației cu ajutorul testului F.	2
LP8. Testarea specificației modelului și simularea variabilelor omise.	2
LP9. Regresie logistică în Python. Codificarea variabilelor binare.	2

Tematica activităților didactice	Numărul de ore
	învățământ cu frecvență
LP10. Modelarea seriilor temporale în Python: AR, MA, ARIMA. ACF și PACF. Testul DF și ADF în Python - implementare și interpretare.	2
Total lucrări practice:	20

8. Referințe bibliografice

Principale	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 4. Wes McKinney, Python for Data Analysis, O'Reilly Media, 3rd Edition, 2022 ISBN: 9781098104009
Suplimentare	5. 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. Utilizarea IA generativă

Permisivitatea 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 appendice 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	Studenții 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				
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.

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 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%