

ANALIZA STATISTICĂ ȘI VIZUALIZAREA DATELOR
1. Date despre disciplină/modul

Facultatea	Calculatoare, Informatică și Microelectronică				
Catedra/departamentul	Informatică și Ingineria Sistemelor				
Ciclul de studii	Studii superioare de licență, ciclul I				
Programul de studiu	0612.3 Știința Datelor				
Anul de studiu	Semestrul	Tip de evaluare	Categoria formativă	Categoria de opționalitate	Credite ECTS
I (învățământ cu frecvență)	4	E	S – unitate de curs de specialitate	A - unitate de curs la alegere	4

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
120	30	15/15	-	30	30

3. Precondiții de acces la disciplină/modul

Conform planului de învățământ	Analiza Matematică, Probabilitate și statistică aplicată, Structuri de date și algoritmi, Matematica discretă, Programarea calculatoarelor, Programare Interactivă.
Conform competențelor	Utilizarea de teorii și instrumente specifice domeniului (algoritmi, metode, tehnici, etc.) pentru analiza algoritmilor de bază din statistică, analiză matematică și ilustrarea datelor grafic.

4. Condiții de desfășurare a procesului educațional pentru

Curs	Pentru prezentarea materialului teoretic în sala de curs este nevoie de tablă, proiector și calculator.
Laborator/seminar	Studentii vor perfecta rapoarte conform condițiilor impuse de indicațiile metodice. Termenul de predare a lucrării de laborator – 2 săptămâni după finalizarea acesteia. Pentru predarea cu întârziere a lucrării aceasta se depunțează cu 1pct./săptămână de întârziere. Pentru desfășurarea seminarelor în auditoriu de studii este nevoie de tablă, proiector și calculator.

5. Competențe specifice acumulate

Competențe profesionale	CP2. Proiectarea și dezvoltarea aplicațiilor K1 Programe/module software adecvate. K2 Componente hardware, instrumente și arhitecturi hardware. K3 Proiectarea funcțională și tehnică. K4 Tehnologiile de ultimă oră. K5 Limbaje de programare. K6 Baze de date (DBMS). K7 Sisteme de operare și platforme software. K8 Mediul de dezvoltare integrat (IDE - integrated development environment). K9 Dezvoltarea rapidă a aplicațiilor. K10 Problemele legate de drepturile de proprietate intelectuală (IPR). K11 Tehnologia de modelare tehnică și limbaje.
-------------------------	---

	<p>K12 Limbajele de definire a interfeței (IDL). K13 Probleme de securitate. CP3. Integrarea componentelor K1 Componente/module hardware/software, indiferent dacă sunt vechi, existente sau noi. K2 Impactul integrării unui sistem asupra organizației sau a sistemului existent. K3 Tehnici de interfațare între module, sisteme și componente. K4 Tehnici de testare a integrării. K5 Instrumentele de dezvoltare (ex. mediul de dezvoltare, gestionare, control al modificărilor și accesul la codul sursă). K6 Bune practici de design</p>
--	---

6. Obiectivele disciplinei/modulului

Obiectivul general	Să descrie conceptele de bază, Să determine rolul vizualizării datelor , Să determine instrumentele și tehnicile utilizate
Obiectivele specifice	<p>Să cunoască măsuri de tendință centrală: medie, mediană, mod precum și măsuri de dispersie: varianță, deviație standard, interval interquartil Să definească noțiunea de regresie, Să poată evalua modelele de regresie Să poată utiliza principii de design pentru vizualizarea datelor, să folosească instrumente de vizualizare: Matplotlib, Seaborn, Plotly și să creeze grafice interactive</p>

7. Conținutul disciplinei/modulului

Tematica activităților didactice	Numărul de ore învățământ cu frecvență
Tematica prelegerilor	
Tema 1. Introducere în Analiza Statistică și Vizualizarea Datelor.	2
Tema 2. Colectarea și Preprocesarea Datelor.	4
Tema 3. Statistici Descriptive	2
Tema 4. Probabilitate și Distribuții de Probabilitate	4
Tema 5. Analiza Regresiei	2
Tema 6. Analiza Factorială.	4
Tema 7. Analiza de Grupare (Clustering)	2
Tema 8. Tehnici de Vizualizare a Datelor.	4
Tema 9. Vizualizarea Datelor Geospațiale.	2
Tema 10. Proiecte de Vizualizare a Datelor.	4
Total prelegeri:	30

Tematica activităților didactice	Numărul de ore învățământ cu frecvență
Tematica lucrărilor de laborator/seminarelor	
Seminar 1. Preprocesarea Datelor.	2
Seminar 2. Statistici Descriptive și Vizualizarea Datelor	4
Seminar 3. Distribuții de Probabilitate și Testare de Ipoteze.	2
Seminar 4. Regresie Liniară	4
Seminar 5. Analiza Clustering	3
Lucrarea de laborator nr 1. Preprocesarea și Explorarea Datelor	2
Lucrarea de laborator nr 2. Vizualizări Statistice de Bază	4
Lucrarea de laborator nr 3. Vizualizarea și Analiza Distribuțiilor de Probabilitate	2

Lucrarea de laborator nr 4. Proiect de Analiză de Regresie	4
Lucrarea de laborator nr 3. Proiect de Vizualizare Interactivă	3
Total lucrări de laborator/seminare:	15/15

8. Referințe bibliografice

Principale	<ol style="list-style-type: none"> Lucia Căbulea, Nicoleta Breaz. INTERPRETAREA STATISTICĂ A INFORMAȚIILOR. ELEMENTE DE DATA MINING ȘI PROGNOZĂ, Universitatea 1 Decembrie 1918” Alba Iulia. http://www.uab.ro/cursuri_perfectionare/program_perfectionare_profesionala/pagini/cursuri/Modul-7-Cabulea-Breaz/curs_modul7.doc T. Andrei, Statistică și econometrie, Ed. Economică, 2003. C. Anghelache, E. Bugudui, S. Gresoi, E. Niculescu, Statistică aplicată-indicatori, sinteze și studii de caz, Ed. Economică, 2006 E.N. Bâzdoacă, S. Matei, N.G. Bâzdoacă, Inițiere în Excel, Ed.Arves, Colecția Inițiere în calculator 2, 2002 Liviu Ciortuz, Alina Munteanu, Elena Badarau Exerciții de învățare automată ed UNIVERSITATII "ALEXANDRU IOAN CUZA" (2015) ISBN: 606-714-197-9 MARCUS, S., NICOLAU, E., STATI, S., 1966. Introducere în lingvistica matematică. București, Editura științifică, p.336. An estimate of an upper bound for the entropy of English. Brown, Della Pietra, Mercer, Della Pietra, Lai. Computational Linguistics, 18(1), pp31-40, 1992. VLAD, A., MITREA, A., MITREA, M., 2005. Limba română scrisă ca sursă de informație. Paideia, România, 286 p. Roman Feldman, James Sauger. The Text Mining Handbook. Advanced Approaches in Analyzing Unstructured Data. Cambridge 2008.
Suplimentare	<ol style="list-style-type: none"> MIHALCEA, R., 2002. Diacritics Restoration: Learning from Letters versus Learning from Words. in Proceedings of the Third International Conference on Intelligent Text Processing and Computational Linguistics, Mexico, p.339-348. Christopher D.Manning, Hinrich Schütze Foundations of Statistical Natural Language Processing Mit Press, 1999. Ronald Rosenfeld. Two decades of Statistical Language Modeling: Where Do We Go From Here? Proceedings of the IEEE, 88(8), 2000. Lillian Lee. "I'm sorry Dave, I'm afraid I can't do that": Linguistics, Statistics, and Natural Language Processing. Computer Science: Reflections on the Field, Reflections from the Field, pp. 111--118, 2004.

9. Evaluare

Forma de învățământ	Periodică		Curentă	Lucrul individual	Examen final
	Atestarea 1	Atestarea 2			
Cu frecvență	15%	15%	15%	15%	40%

Standard minim de performanță

Prezența și activitatea la prelegeri, seminare 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 algoritmilor de bază din teoria limbajelor formale, automatelor finite și automatelor pushdown.