

S.A.104 ANALYSE STATISTIQUE ET VISUALISATION DES DONNÉES

1. Données sur la discipline/le module

Faculté	Ordinateurs, informatique et microélectronique				
Chaire/département	Informatique et ingénierie des systèmes				
Cycle d'études	Licence, premier cycle				
Programme d'études	0612.3 Science des données				
Année d'études	semestre	Type d'évaluation	Catégorie formative	Catégorie d'optionnalité	Crédits ECTS
I (enseignement à temps partiel)	4	C'est	S – unité de cours spécialisée	A - unité de cours au choix	4

2. Durée totale estimée

Nombre total d'heures dans le programme	Dont				
	Horaires de l'auditorium		Travail individuel		
	cours	Laboratoire/séminaire	Projet de l'année	Étude de la matière théorique	Préparation de la candidature
120	30	15/15	-	30	30

3. Prérequis pour l'accès à la discipline/module

Selon le programme	Analyse mathématique, probabilités et statistiques appliquées, structures de données et algorithmes, mathématiques discrètes, programmation informatique, programmation interactive.
Selon les compétences	Utilisation de théories et d'outils spécifiques au domaine (algorithmes, méthodes, techniques, etc.) pour l'analyse des algorithmes de base en statistique, analyse mathématique et illustration graphique de données.

4. Conditions de réalisation du processus éducatif pour

Cours	Pour présenter le matériel théorique en classe, un tableau noir, un projecteur et un ordinateur sont nécessaires.
Laboratoire / séminaire	Les étudiants rédigeront des rapports selon les conditions imposées par les instructions méthodologiques. La date limite de remise des travaux de laboratoire est de 2 semaines après leur achèvement. En cas de remise tardive du travail, il sera déduit 1 point/semaine de retard. Pour mener des séminaires dans l'auditorium d'étude, un tableau blanc, un projecteur et un ordinateur sont nécessaires.

5. Compétences spécifiques acquises

Compétences professionnelles	<p>CP2. Conception et développement d'applications</p> <p>K1 Logiciels/modules appropriés. Composants matériels, outils et architectures matérielles K2. K3 Conception fonctionnelle et technique. K4 Technologies de pointe. Langages de programmation K5. K6 (SGBD). K7 Systèmes d'exploitation et plateformes logicielles. Environnement de développement intégré (IDE) K8. K9 Développement rapide d'applications. Questions relatives aux droits de propriété intellectuelle (DPI) du K10.</p>
------------------------------	---

	<p>K11 Modélisation technique, technologie et langages. Langages de définition d'interface K12 (IDL). Problèmes de sécurité K13.</p> <p>CP3. Intégration des composants</p> <p>K1 Composants/modules matériels/logiciels, qu'ils soient anciens, existants ou nouveaux. K2 L'impact de l'intégration d'un système sur l'organisation ou le système existant. K3 Techniques d'interfaçage entre modules, systèmes et composants. Techniques de test d'intégration K4. K5 Outils de développement (par exemple, environnement de développement, gestion, contrôle des modifications et accès au code source). K6 Bonnes pratiques de conception</p>
--	--

6. Objectifs du cours/module

Objectif général	Décrire les concepts de base, Déterminer le rôle de la visualisation des données, Déterminer les outils et techniques utilisés
Objectifs spécifiques	<p>Connaître les mesures de tendance centrale : moyenne, médiane, mode ainsi que les mesures de dispersion : variance, écart type, intervalle interquartile</p> <p>Définir la notion de régression, Être capable d'évaluer des modèles de régression</p> <p>Être capable d'utiliser les principes de conception pour la visualisation des données, utiliser les outils de visualisation : Matplotlib, Seaborn, Plotly et créer des graphiques interactifs</p>

7. Contenu du cours/module

Thème des activités pédagogiques	Nombre d'heures
	éducation à temps partiel
Sujets de cours	
Thème 1. Introduction à l'analyse statistique et à la visualisation des données.	2
Thème 2. Collecte et prétraitement des données.	4
Thème 3. Statistiques descriptives	2
Thème 4. Probabilité et distributions de probabilité	4
Thème 5. Analyse de régression	2
Thème 6. Analyse factorielle.	4
Thème 7. Analyse de clustering	2
Thème 8. Techniques de visualisation des données.	4
Devoirs 9 . Visualisation des données géospatiales.	2
Devoirs 10 . Projets de visualisation de données .	4
Nombre total de conférences :	30

Thème des activités pédagogiques	Nombre d'heures
	éducation à temps partiel
Thèmes de travaux de laboratoire/séminaire	
Séminaire 1. Prétraitement des données.	2
Séminaire 2. Statistiques descriptives et visualisation des données	4
Séminaire 3. Distributions de probabilité et tests d'hypothèses.	2
Séminaire 4. Régression linéaire	4
Séminaire 5. Analyse de clustering	3

Travaux pratiques n° 1. Prétraitement et exploration des données	2
Travaux pratiques n° 2. Visualisations statistiques de base	4
Travaux pratiques n° 3. Visualisation et analyse des distributions de probabilité	2
Travaux pratiques n° 4. Projet d'analyse de régression	4
Travaux pratiques n° 3. Projet de visualisation interactive	3
Total des travaux de laboratoire/séminaire :	15/15

8. Références bibliographiques

principal	<ol style="list-style-type: none"> 1. Lucie Cabulea, Nicoleta Breaz. INTERPRÉTATION STATISTIQUE DE L'INFORMATION. ÉLÉMENTS D'EXPLOSION DE DONNÉES ET DE PRÉVISION, Université 1er décembre 1918 » Alba Iulia. 2. http://www.uab.ro/cursuri_perfectionare/program_perfectionare_profesionala/pagini/cursuri/Modul-7-Cabulea-Breaz/curs_modul7.doc 3. T. Andrei, Statistique et économétrie, éd. Économique, 2003. 4. C. Anghelache, E. Bugudui, S. Gresoi, E. Niculescu, Statistiques appliquées - Indicateurs, synthèses et études de cas, éd. Économique, 2006 5. FR Bâzdoacă, S. Matei, NG Bâzdoacă, Introduction à Excel, Ed.Arves, Computer Introduction Collection 2, 2002 6. Liviu Ciortuz, Alina Munteanu, Elena Badarau Exercices d'apprentissage automatique publiés par l'UNIVERSITÉ « ALEXANDRU IOAN CUZA » (2015) ISBN : 606-714-197-9 7. MARCUS, S., NICOLAU, E., STATI, S., 1966. Introduction à la linguistique mathématique. Bucarest, Maison d'édition scientifique, p.336. 8. Une estimation d'une limite supérieure pour l'entropie de l'anglais. Brown, Della Pietra, Mercer, Della Pietra, Lai. Linguistique computationnelle, 18(1), pp31-40, 1992. 9. VLAD, A., MITREA, A., MITREA, M., 2005. La langue roumaine écrite comme source d'information. Paideia, Roumanie, 286 p. 10. Roman Feldman, James Sauger. Le manuel d'exploration de texte. Approches avancées dans l'analyse des données non structurées. Cambridge 2008.
plus	<ol style="list-style-type: none"> 1. MIHALCEA, R., 2002. Restauration des signes diacritiques : apprentissage à partir des lettres versus apprentissage à partir des mots. dans les actes de la troisième conférence internationale sur le traitement intelligent de texte et la linguistique informatique, Mexique, pp.339-348. 2. Christopher D. Manning, Hinrich Schutze Fondements du traitement statistique du langage naturel Mit Press, 1999. 3. Ronald Rosenfeld. Deux décennies de modélisation statistique du langage : où allons-nous à partir de maintenant ? Actes de l'IEEE, 88(8), 2000. 4. Lilian Lee. « Je suis désolé Dave, j'ai bien peur de ne pas pouvoir faire ça » : Linguistique, statistiques et traitement du langage naturel. Informatique : Réflexions sur le terrain, Réflexions du terrain, pp. 111-118, 2004.

9. Utilisation de l'IA générative

Autorisation d'utilisation	<p>L'utilisation de l'IA générative dans les devoirs et les projets est autorisée, à condition que les étudiants respectent les règles suivantes :</p> <ul style="list-style-type: none"> • L'IA générative peut être utilisée pour générer des idées, des structures de texte ou du code, mais tous les matériaux générés doivent être examinés et ajustés par l'étudiant pour garantir qu'ils répondent aux exigences académiques. • Toute utilisation de l'IA générative doit être déclarée dans la section annexe de chaque article, en utilisant la phrase : « Lors de la préparation de cet article, l'auteur a utilisé [NOM DE L'OUTIL/SERVICE] aux fins de [RAISON]. Après avoir utilisé cet
-----------------------------------	--

	outil/service, l'auteur a révisé et modifié le contenu si nécessaire et assume l'entière responsabilité du contenu de l'article . »
Restrictions d'utilisation	<p>Les étudiants ne devraient pas considérer l'IA générative comme une source d'information fiable car elle ne fournit pas de références claires ni de sources documentées.</p> <ul style="list-style-type: none"> • La citation directe du contenu généré par l'IA dans les articles universitaires comme s'il s'agissait d'une source primaire n' est pas autorisée . • Les activités dans lesquelles l'utilisation de l'IA générative est interdite sont spécifiées par l'enseignant et sont généralement des évaluations intermédiaires et finales ou qui n'impliquent pas d'activités de développement des compétences professionnelles .

10. ÉVALUATION

Formulaire d'éducation	périodique		flux	Travail individuel	Examen final
	Certification 1	Certification 2			
Avec fréquence	15%	15%	15%	15%	40%
Norme de performance minimale					
<p>Présence et activité aux cours magistraux, séminaires et travaux de laboratoire ; Obtenir une note minimale de « 5 » dans chacune des attestations et travaux de laboratoire; Démonstration dans l'épreuve finale de la connaissance des algorithmes de base dans la théorie des langages formels, des automates finis et des automates à pile.</p>					