

	Fiche module		Mise à jour :
	Cycle de formation d'ingénieurs en Informatique		Page 1 sur 4

Module : Traitement analytique avec Python				Code	
				ING-4-SDIA-S8-P1	
Période	S8- P1	Volume horaire	42h	ECTS	4

Responsable	Mr. Nadjib Bendaoud	email	nadjib.bendaoud@tek-up.tn
Equipe pédagogique	Mr. Nadjib Bendaoud		

1. Objectifs de Module (*Savoirs, aptitudes et compétences*)

Ce module porte sur les algorithmes de Machine learning, traitement et l'analyse des données.

Acquis d'apprentissage :

A la fin de cet enseignement, l'élève sera capable de :

- Maîtriser les différents algorithmes de Machine Learning. **(C1.2)**
- Visualisation des données. **(C1.3)**
- Caractériser le traitement des données. **(C1.4)**
- Maîtriser module Machine Learning Sklearn. **(1.4)**

2. Pré-requis(*autres UE et compétences indispensables pour suivre l'UE concernée*)

1. Python
2. Machine Learning
3. Natural Language programming

3. Répartition d'Horaire de Module

Intitulé de l'élément d'enseignement	Total	Cours	TD	Atelier	PR
Module : Traitement analytique avec Python			6	36	

4. Méthodes pédagogiques et moyens spécifiques au Module

(*pédagogie d'enseignement, ouvrages de références, outils matériels et logiciels*)

- Supports de Cours
- Projecteur et Tableau
- Travaux pratiques

Bibliographie		
Titre	Auteur(s)	Edition

	Fiche module		Mise à jour :
	Cycle de formation d'ingénieurs en Informatique		Page 2 sur 4

5. Contenu (Descriptifs et plans des cours / Déroulement / Détail de l'évaluation de l'activité pratique)		Durée allouée	
Module 1: Machine Learning			
Séance 1 : La visualization des données <ul style="list-style-type: none">MatplotlibPandas visualizationSeabornggplot		Atelier	3H
Séance 2 : Applications des Classifications <ul style="list-style-type: none">Classification binaire : base des données des diabètesEvaluations de la performance des modèles en utilisant : l'accuracy/ recall/ precision AUc, f1-score/ roc-curveAnalyse des résultats		Atelier	3 H
Séance 3 : Applications des Classifications <ul style="list-style-type: none">Mono-classe : base des données des diabètesClassification multi-classe : base irisEvaluations de la performance des modèles en utilisant : l'accuracy/ recall/ precision AUc, f1-score/ roc-curveAnalyse des résultats		Atelier	3H
Séance 4 : Applications des Régressions <ul style="list-style-type: none">Projet : Prédiction du prix des maisonsLes algorithmes : La régression linéaire simple et multipleEvaluations de la performance des modèles en utilisant : explained variance, RMSE, MSE,...		Atelier	3H
Séance 5 : Applications des Régressions <ul style="list-style-type: none">Projet : Prédiction du prix des maisonsLes algorithmes : décision tree/XGboost/ random forestEvaluations de la performance des modèles en utilisant : explained variance, RMSE, MSE,...		Atelier	3H
Séance 6 : Application de Réduction de Dimension <ul style="list-style-type: none">Projet : la reconnaissance facialeBase des données : YalefacesAlgorithmes : PCA pour la réduction des dimensionsSVM pour la classificationAnalyse des résultats		Atelier	3H
Séance 7 : <ul style="list-style-type: none">Activités (TD)		TD	3H
Séance 8 : Application de Réduction de Dimension <ul style="list-style-type: none">Projet : la reconnaissance facialeBase des données : YalefacesAlgorithmes : LDA pour la réduction des dimensionsLDA pour la classificationAnalyse des résultats		Atelier	3H
Séance 9 : Applications de la Classification non supervisé et clustering		Atelier	3H

	Fiche module		Mise à jour :
	Cycle de formation d'ingénieurs en Informatique		Page 3 sur 4

<ul style="list-style-type: none"> • Générer des données artificielles • Les algorithmes utilisés pour chercher nombre des classes : <ul style="list-style-type: none"> - Elbow method - Silhouette score, - validity index, • Les algorithmes utilisés pour déterminer les étiquettes : <ul style="list-style-type: none"> - Kmeans - Fuzzy Kmeans • Analyse des résultats 		
Séance 10 : Applications de la Classification non supervisé et clustering <ul style="list-style-type: none"> • Générer des données artificielles • Les algorithmes utilisés pour chercher nombre des classes : <ul style="list-style-type: none"> - hierarchical clustering, - DBSCAN • Les algorithmes utilisés pour déterminer les étiquettes : <ul style="list-style-type: none"> - Kmeans • Analyse des résultats 	Atelier	3H
Séance 11 : Projet Classification des textes <ul style="list-style-type: none"> • Base des données : Spam.csv • Algorithmes : SVM, logistic regression, decision tree, Xgboost, random forest • NLP pour le nettoyage des données 	Atelier	3 H
Séance 12 : Continuité du Projet Classification des textes <ul style="list-style-type: none"> • Base des données : Spam.csv • Algorithmes : SVM, logistic regression, decision tree, Xgboost, random forest • NLP pour le nettoyage des données 	Atelier	3 H
Séance 13 : Chatbot <ul style="list-style-type: none"> • Chatbot qui répond aux questions Data Science Students 	Atelier	3H
Séance 14 : <ul style="list-style-type: none"> • Activités (TD) 	TD	3H


6. Mode d'évaluation de Module (nombre, types et pondération des contrôles)

Eléments d'enseignement	Coeff	DS	EX	TP	PR
Module – Traitements analytiques avec Python	2		60%	40%	

Pour valider le module, les étudiants passeront un examen dont le coefficient est de 60%, et un TP dont le coefficient est de 40%.

La durée de tous les examens (Examen, TP...) est de 1h30.

Quant à l'examen, il est planifié après l'écoulement des 14 semaines et portera sur toutes les thématiques enseignées tout au long des 42 heures.

	Fiche module		Mise à jour :
	Cycle de formation d'ingénieurs en Informatique		Page 4 sur 4

Concernant le TP, il est planifié une semaine avant l'examen et testera les connaissances acquises tout au long du module.

Le module est validé si l'étudiant obtient une moyenne supérieure ou égal à 10 sur 20.