

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

<i>Responsable</i>	<b>Mr. Nadjib Bendaoud</b>	<i>email</i>	nadjib.bendaoud@tek-up.tn
<i>Equipe pédagogique</i>	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

<i>Intitulé de l'élément d'enseignement</i>	<i>Total</i>	<i>Cours</i>	<i>TD</i>	<i>Atelier</i>	<i>PR</i>
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

<b>Titre</b>	<b>Auteur(s)</b>	<b>Edition</b>

<b>5. Contenu</b> ( <i>Descriptifs et plans des cours / Déroulement / Détail de l'évaluation de l'activité pratique</i> )	Durée allouée	
<b>Module 1: Machine Learning</b>		
<b>Séance 1 : La visualization des données</b> <ul style="list-style-type: none"> <li>• Matplotlib</li> <li>• Pandas visualization</li> <li>• Seaborn</li> <li>• ggpplot</li> </ul>	Atelier	3H
<b>Séance 2 : Applications des Classifications</b> <ul style="list-style-type: none"> <li>• Classification binaire : base des données des diabètes</li> <li>• Evaluations de la performance des modèles en utilisant : l'accuracy/ recall/ precision AUC, f1-score/ roc-curve</li> <li>• Analyse des résultats</li> </ul>	Atelier	3 H
<b>Séance 3 : Applications des Classifications</b> <ul style="list-style-type: none"> <li>• Mono-classe : base des données des diabètes</li> <li>• Classification multi-classe : base iris</li> <li>• Evaluations de la performance des modèles en utilisant : l'accuracy/ recall/ precision AUC, f1-score/ roc-curve</li> <li>• Analyse des résultats</li> </ul>	Atelier	3H
<b>Séance 4 : Applications des Régressions</b> <ul style="list-style-type: none"> <li>• Projet : Prédiction du prix des maisons</li> <li>• Les algorithmes : La régression linéaire simple et multiple</li> </ul> Evaluations de la performance des modèles en utilisant : explained variance, RMSE, MSE,..	Atelier	3H
<b>Séance 5 : Applications des Régressions</b> <ul style="list-style-type: none"> <li>• Projet : Prédiction du prix des maisons</li> <li>• Les algorithmes : décision tree/XGboost/ random forest</li> <li>• Evaluations de la performance des modèles en utilisant : explained variance, RMSE, MSE,..</li> </ul>	Atelier	3H
<b>Séance 6 : Application de Réduction de Dimension</b> <ul style="list-style-type: none"> <li>• Projet : la reconnaissance faciale</li> <li>• Base des données : Yalefaces</li> <li>• Algorithmes : PCA pour la réduction des dimensions</li> <li>• SVM pour la classification</li> <li>• Analyse des résultats</li> </ul>	Atelier	3H
<b>Séance 7 :</b> <ul style="list-style-type: none"> <li>• Activités (TD)</li> </ul>	TD	3H
<b>Séance 8 : Application de Réduction de Dimension</b> <ul style="list-style-type: none"> <li>• Projet : la reconnaissance faciale</li> <li>• Base des données : Yalefaces</li> <li>• Algorithmes : LDA pour la réduction des dimensions</li> <li>• LDA pour la classification</li> <li>• Analyse des résultats</li> </ul>	Atelier	3H
<b>Séance 9 : Applications de la Classification non supervisé et clustering</b>	Atelier	3H

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

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.