

SUJETS POSSIBLES POUR L'UE (

date de soutenance 14 et 16 mai 2024

)

Introduction à Pandas et NumPy : Présentation des structures de données, manipulation et nettoyage de données avec Pandas et NumPy.

Visualisation de données avec Matplotlib et Seaborn : Techniques de base et avancées pour visualiser des ensembles de données complexes.

Analyse exploratoire de données (EDA) : Méthodes et outils pour explorer des ensembles de données, détecter des anomalies et des tendances.

Traitement des données manquantes : Stratégies pour gérer les valeurs NaN dans les ensembles de données.

Web scraping avec Python : Utilisation de BeautifulSoup ou Scrapy pour collecter des données à partir de pages web.

Analyse de séries temporelles : Techniques pour analyser les données temporelles, y compris la prévision.

Machine Learning avec Scikit-Learn : Introduction aux concepts de base du machine learning et application à l'aide de la bibliothèque Scikit-Learn.

Traitement du langage naturel (NLP) avec NLTK ou spaCy : Fondamentaux du NLP et mise en œuvre de projets simples de traitement du texte.

Analyse de réseaux sociaux : Utilisation de bibliothèques comme NetworkX pour analyser les réseaux sociaux.

Création de dashboards interactifs avec Dash et Plotly : Développement d'applications web pour la visualisation interactive de données.

Analyse de données géospatiales : Manipulation et visualisation de données géospatiales avec GeoPandas et autres outils spécifiques.

Optimisation de code pour l'analyse de données : Techniques pour améliorer la performance du code Python lors du traitement de grandes ensembles de données.

Utilisation de l'API pour la collecte de données : Écriture de scripts pour extraire des données à partir d'APIs web.

Deep Learning avec TensorFlow ou PyTorch : Introduction aux concepts du deep learning et mise en œuvre d'un projet simple.

Analyse de données financières : Utilisation de Python pour analyser des données de marché, y compris des stratégies de trading algorithmique.

Détection d'anomalies dans les ensembles de données : Techniques et outils pour identifier les données aberrantes.

Gestion de bases de données avec SQLAlchemy : Introduction à l'interaction avec des bases de données relationnelles en Python.

Visualisation de données 3D avec Matplotlib et Plotly : Techniques pour créer des graphiques tridimensionnels pour une meilleure analyse des données.

Analyse comparative de bibliothèques Python pour l'analyse de données : Évaluation des forces et faiblesses de différentes bibliothèques (Pandas vs. Dask, Matplotlib vs. Seaborn, etc.).

Projets d'analyse de données réels : Application des compétences acquises sur des ensembles de données réels et présentation des résultats.