République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Direction Générale des Enseignements et de la formation supérieurs

**Programmes transversaux de
programmation informatique et
d'intelligence artificielle pour le cycle de
Master en Sciences Humaines et Sociales**

Janvier 2025

Programmation informatique appliquée aux Sciences Humaines et Sociales

Compétences visées :

- Utilisation des outils informatiques pour l'acquisition, le traitement, la production et la diffusion de l'information essentiellement pour les sciences sociales et humaines.
- Analyse des données d'enquête en utilisant les technologies informatiques.
- Interprétation des résultats statistiques sur des corpus de textes.

Objectifs :

- Maîtrise des méthodes pour l'analyse statistiques en Sciences Sociales et Humaines.
- Aptitude à utiliser des logiciels : R, SPSS, Python, Excel avancé, JMP (avec licence).
- Aptitude à construire et traiter un corpus de textes à partir du Web.

Méthode pédagogique

- Cours magistraux.
- Travaux dirigés.
- Travaux pratiques.
- Interactivité maximale enseignant – étudiants.
- Travail personnel (surtout pour la programmation).
- Evaluation (identique aux textes en vigueur).
- Tables rondes (la sécurité de données, impact de l'intelligence artificielle dans les sciences sociales et humaines, etc.).

Répartition horaire :

- 2^{ème} semestre de Master (M1)
- Durée (12 semaines):
 - o Cours magistraux (1h30/semaine).
 - o Travaux dirigés (1h30/semaine).
 - o Travaux pratiques (3h/semaine)

Programme :

Chapitre 1. Connaissances informatiques et traitement informatique textuel (1 semaine)

- Logiciel et matériel,
- Principes de la programmation,
- Utilisation d'internet,
- Gestion de bases de données,
- Principes d'extraction et analyse de corpus textuels.

Chapitre 2 : Maîtrise des méthodes pour l'analyse statistiques en sciences sociales (1 semaine)

- Les variables,
- Moyenne,
- Ecart type et variance,
- Analyse de la variance,

- Loi de probabilités usuelles,
- Test de Khi-deux,
- Coefficient de corrélation,
- Analyse linéaire simple,
- Analyse linéaire multiple,
- Introduction à l'analyse factorielle.

Chapitre 3 : Logiciels et outils pour le traitement de textes (5 semaines)

- **Apprentissage pratique obligatoire :**
 - R,
 - SPSS application aux enquêtes,
 - Python : analyse de textes (Bibliothèques en Python : NLTK, spaCy, gensim, pandas, etc.),
 - Programmation en Python :
 - Boucles et conditions
 - Fichiers,
 - Librairie de base
 - Ecriture de scripts pour :
 - traiter des fichiers textes (recherche, tri)
 - gérer des rapports simples (PDF, Excel)
 - Excel :
 - Fonctions avancées,
 - Les macros,
 - Application aux enquêtes,
 - Analyses de base,
 - Programmation en Excel de la régression linéaire,
 - Tableaux croisés et Graphiques :
 - Tableaux croisés,
 - Histogrammes,
 - Diagrammes en barres,
 - Araignée,
 - Etc.
- **Apprentissage optionnel** (Obligatoire si Acquisition de la licence) :
 - Utilisation du JMP,

Chapitre 4 : Extraction de textes et Analyse de données textuelles (3 semaines)

- Construction d'un corpus de textes :
 - Types de corpus textuels : discours, articles, tweets...,
 - Applications : extraction d'informations, clustering,
 - Approche par scraping, API.
- Les créateurs de contenus et leur rôle dans la production et la diffusion de l'information.
- Les questionnaires, les questions ouvertes et les entretiens.

Chapitre 5 : Normalisation et vectorisation de Textes (2 semaines)

- **Standardisation de textes :**
 - Standardisation de textes :
 - Suppression de la ponctuation, des chiffres, des caractères spéciaux.
 - Conversion en minuscules, suppression des espaces blancs superflus.
 - Tokenisation : découper un texte en mots ou phrases.
 - Applications.

- **Normalisation de textes :**
 - Suppression des mots vides : stopwords ("le", "et", "à").
 - Stemming (Lemmatization) : réduire les mots à leur racine ou forme canonique (utilisation d'un algorithme).
 - Détection et gestion des fautes d'orthographe (optionnel).
 - Application.

- **Vectorisation de textes :**
 - Sac de mots (Bag of Words) : Représentation matricielle du texte.
 - TF-IDF : Term Frequency-Inverse Document Frequency pour extraire l'importance des mots dans un corpus.
 - Word Embeddings : Introduction aux représentations vectorielles des mots (Algorithmes Word2Vec, GloVe).
 - Applications.

Travaux dirigés :

- Chapitre 1 :
 - Exercices simples de programmation,
 - Vue d'un SGBD du type ACCESS.
- Chapitre 2 :
 - Applications manuelles sur les méthodes statistiques (moyenne, écart-type, variance...) sur des cas SHS,
 - Application d'analyse factorielle.
- Chapitre 3 :
 - Exercices de Python.
 - Exercices sur les fonctions et les macros d'Excel.
 - Tableaux comparatif des langages R, SPSS, Python faisant apparaître les avantages et les inconvénients,
 - Approfondir la notion de corpus de textes en SHS,
- Chapitre 4 :
 - Notions d'extraction de corpus de textes à partir de sites Web, email, etc :
 - Structuration d'un site Web, emails et toute autre source d'information
 - Approche d'extraction par scraping,
 - Approche d'extraction par API.

- Rôle de la question ouverte et les entretiens dans la construction de corpus.
- Chapitre 5 :
 - Exercices sur la tokenisation, lemmatisation.
 - Matrices termes – documents
 - Vectorisation des termes
 - Applications TF/IDF
 - Analyse

Travaux pratiques :

- Chapitre 1 :
 - Les algorithmes,
 - Exemples de programmation en utilisant un langage algorithmique,
 - Traduction des algorithmes en langage de programmation (C par exemple),
 - Exercices pratiques sur Internet,
 - Exercices sur le SGBD Microsoft Access.
- Chapitre 2 :
 - Application manuelle des méthodes statistiques aux textes (chapitre 2) sur un échantillon réel,
- Chapitre 3 :
 - Exercices de Python.
 - Analyse sur des enquêtes sociologiques en utilisant la technologie informatique.
 - Importation et gestion de corpus textuels dans Python, R et SPSS.
 - Applications informatiques sur les méthodes statistiques (interprétation des résultats) sur des échantillons réels (à télécharger) et en utilisant les langages de programmation étudiés : R, SPSS, Python,
 - Application d'Excel sur des textes, numériques, graphismes, tableaux croisés.
- Chapitre 4 :
 - Applications d'extraction de corpus de textes à partir de sites Web, email, etc :
 - Approche d'extraction par scraping,
 - Approche d'extraction par API.
 - Application d'une enquête avec question ouverte et des entretiens.
 - Application à la normalisation de textes.
 - Application à la vectorisation de textes.
- Chapitre 5 :
 - Types de corpus textuels : discours, articles, tweets...,
 - Applications : extraction d'informations, clustering,
 - Application des approches scraping et API pour l'extraction de textes (ou de mots, ou d'emails,...) à partir de Sites Web (ou autres).

Bibliographie

- Pour le chapitre 1 :
 - Ouvrages : Introduction à l'informatique, Principes de la programmation, Gestion de bases de données,
 - Duteil-Mougel Carine (2005) : Les mécanismes persuasifs des textes politiques, Corpus N°4. Lien : <http://corpus.revues.org/357>
 - J. Guilhaumou (2002) : Le corpus en analyse de discours, perspectives historiques.. Corpus N°1. Lien : <http://corpus.revues.org/8>
- Pour le chapitre 2 :
 - Ouvrages classiques sur les variables quantitatives et qualitatives.
- Pour le chapitre 3 :
 - E.Schultz et M.Bussonnier (2020) : Python pour les SHS. Introduction à la programmation de données. Presses Universitaires de Rennes.
 - C.Paroissin, (2021) : Pratique de la data science avec R : arranger, visualiser, analyser et présenter des données. Paris : Ellipses, DL 2021.
- Chapitre 5 :
 - Baddari Kamel et Herzallah Abdelkarim (2015) : La recherche de l'information, livre – OPU -Algérie.
- Chapitre 6 :
 - S.Balech et C.Benavent : NLP texte minig V4.0, (Paris Dauphine – 12/2019) : lien : https://www.researchgate.net/publication/337744581_NLP_text_mining_V4_0_-_une_introduction_-_cours_programme_doctoral

Intelligence artificielle appliquée aux Sciences Humaines et Sociales

Compétences visées :

- Identifier les opportunités de l'intelligence artificielle en sciences humaines et sociales.
- Comprendre quelles peuvent être les différentes applications de l'IA pour les SHS.
- Comprendre les implications éthiques de l'IA et les bonnes pratiques de son utilisation.
- Capacité de réaliser le Clustering de comportements sociaux dans des enquêtes et modélisation prédictive.
- Comprendre les sentiments sur les réseaux et comment les analyser.
- Capacité d'extraire des textes dans des discours politiques.
- Capacité d'analyser les réseaux sociologiques.

Objectifs :

- Maîtrise des algorithmes IA pour le traitement textuel des enquêtes
- Aptitude à utiliser le logiciel Python,
- Comprendre les approches de l'IA dans la résolution de problèmes,
- Aptitude à traiter un corpus de textes.

Méthode pédagogique

- Cours magistraux.
- Travaux dirigés.
- Travaux pratiques.
- Mini-projet (réaliser au moins 3 sujets parmi les 5 proposés)
- Interactivité maximale enseignant – étudiants.
- Travail personnel (surtout pour la programmation).
- Evaluation (identique aux textes en vigueur).
- Tables rondes (Genèse et apparition de l'IA, Impact de l'IA sur la réflexion et les méthodologies en SHS, IA vecteur de nouvelles problématiques scientifiques et sociétales, Perspectives de l'IA, etc.).

Répartition horaire :

- 1^{er} semestre de Master (M2)
- Durée 12 semaines :
 - o Cours magistraux (1h30/semaine).
 - o Travaux dirigés (1h30/semaine).
 - o Travaux pratiques (3/semaine)
- Prérequis : avoir franchi le module : Informatique appliquée à l'analyse statistique de textes en Sciences Sociales et Humaines (Semestre 2 du M1)

Programme

Chapitre 1 : Introduction

- **Introduction à l'Intelligence artificielle (IA) (1 semaine, si insuffisant utiliser 1 séance de TP pour compléter)**
 - o Survol historique de l'intelligence artificielle,
 - o IA étroite (ou faible)
 - o IA générale (ou forte)
 - o Apprentissage supervisé et apprentissage non supervisé
 - o Les approches de l'IA dans la résolution de problèmes,
 - o Perspectives.

Chapitre 2 : Rappel et Complément sur le langage (2 semaines)

- o Les fondamentaux
- o Les bibliothèques Pandas, Numpy, etc.
- o Introduction à la programmation objet,
- o Lire et traiter des fichiers textes (recherche, tri)
- o gérer des rapports simples (PDF, Excel)

Chapitre 3 : Types d'IA et Approches (6 semaines)

- **IA Symbolique (système expert)**
- **Apprentissage Automatique (Machine Learning)**
 - o Machine Learning (ML), avec les exemples comme : Classification d'emails comme "spam" ou "non-spam", recommandations, identification d'images (ex : des visages), etc.
- **Apprentissage Profond (Deep Learning)**
 - o Deep Learning
 - o Génération de texte avec Transformers (ou autre).
- **IA par Renforcement (Renforcement Learning)**
 - o Apprentissage par renforcement (RL)

Mini-projet (3 semaines – travail personnel encadré – à distribuer à la fin du chapitre 2) :

Traitez 1 sujet au choix parmi :

Sujet 1. Analyse de Sentiments sur les Réseaux Sociaux

Objectif :

Analyser des corpus de texte provenant des réseaux sociaux (Twitter, Facebook, etc.) pour déterminer les sentiments des utilisateurs à propos d'un sujet particulier (ex. : une campagne politique, une crise sanitaire, ou un événement social).

Outils : Python, TextBlob, VADER, Tweepy (pour Twitter), pandas, matplotlib, GeoPandas.

Sujet 2. Clustering des Comportements Sociaux dans des Enquêtes

Objectif :

Utiliser des techniques de clustering pour regrouper des individus ayant des comportements sociaux similaires à partir de données d'enquête. Cela peut être utile pour segmenter une population en groupes d'intérêt distincts.

Outils : Python, pandas, sklearn, matplotlib, seaborn, scikit-learn.

Sujet 3. Analyse de Texte pour l'Extraction d'Informations dans des Discours Politiques

Objectif :

- Utiliser l'IA pour extraire des thèmes principaux et des entités nommées (noms de personnes, lieux, dates) à partir de discours politiques. Cela peut être utilisé pour comprendre les préoccupations des politiciens et analyser leur rhétorique.

Outils : Python, spaCy, Transformers (BERT), gensim (LDA), matplotlib, WordCloud.

Sujet 4. Modélisation Prédictive des Comportements Sociaux à partir de Données de Consommation

Objectif :

Construire un modèle prédictif pour estimer les comportements sociaux des consommateurs (par exemple, la probabilité d'achat de produits ou de participation à des événements sociaux) en fonction de diverses variables.

Outils : Python, sklearn, pandas, matplotlib, seaborn.

Sujet 5. Analyse des Réseaux Sociaux (Sociologie des Réseaux)

Objectif :

Utiliser des techniques d'analyse des réseaux sociaux (Social Network Analysis - SNA) pour étudier les relations entre individus dans un réseau social, en utilisant des données structurées.

Outils : Python, NetworkX, Gephi, matplotlib, pandas.

Travaux dirigés / Travaux pratiques

Chapitre 1 :

1. Reprendre les notions vues dans le cours (chapitre 1) et approfondir la notion d'apprentissage supervisé et non supervisé
2. Quelle est la différence entre apprentissage supervisé et apprentissage non supervisé ?

3. Donner des exemples d'apprentissage
4. Cas à faire en travail personnel :
 Etudier la littérature sur l'IA afin de déterminer si les ordinateurs actuels peuvent mener à bien les opérations suivantes :
 - Jouer correctement au ping-pong
 - Conduire une voiture au centre d'une ville
 - Découvrir et démontrer des théorèmes mathématiques
 - Ecrire une histoire drôle
 - Donner un avis pertinent dans un domaine juridique donné
 - Réaliser une opération chirurgicale complexe

Chapitre 2 :

- Exercices en Python : programmes, les listes et les tuples, manipulation de données avec les bibliothèques vues au cours.

Chapitre 3 :

1. Exercice sur le système expert :
 - Soit les règles suivants en chainage avant :
 - R1 : $A \rightarrow E$
 - R2 : $B \rightarrow D$
 - R3 : $H \rightarrow A \cap F$
 - R4 : $E \cap G \rightarrow C$
 - R5 : $E \cap K \rightarrow B$
 - R6 : $D \cap E \cap K \rightarrow C$
 - R7 : $G \cap K \cap F \rightarrow A$

Ordre des règles : R3 R7 R1 R4 R5 R2 R6

Faits non déductibles : H, G, K

Base de faits initiale : H, K

But : C

Questions :

Faire le graphe correspondant aux règles

Dérouler le système en chaînage avant.

2. Introduire quelques algorithmes utilisés dans le ML :
 - a. Régression linéaire,
 - b. Gradient descendant
 - c. Classification en catégories basées sur les attributs
 - d. Regroupement d'objets similaires en clusters
 - e. Etc.
3. Donner des exemples sur la classification, les recommandations, l'identification d'images.
4. Exercices sur chaque type d'apprentissage :
 - a. ML
 - b. Deep learning (CNN, RNN)
 - c. renforcement

Travaux pratiques

Les séances de TP seront réservées pour la programmation des exercices vus en TD en utilisant Python et à la réalisation d'un mini-projet.

Bibliographie :

- R.Inglehart (1997) : Modernization and Postmodernization : Cultural , economic, and political chnage in 43 societies – Princeton University Press.
- Western Values are Diverging from the rest of the world's. – The Economiste – 2023.
- Western Michigan University : World Values Survey. Lien : <https://wmich.edu/economics/world-values-survey>
- Ganascia, J.Gabriel (2024) : l'IA expliquée aux humains. Paris France- Edition le Seuil.
- Anglais, Lise, Dilhac, Antione, Dratwa, Jim et al. (2023) : L'éthique au cœur de l'IA. Quebec Obvia.
- J.Robert (2024) : Natural Language Processing (NLP) : définition et principes – Datasciences. Lien : <https://datascientest.com/introduction-au-nlp-natural-language-processing>
- Qu'est-ce que le traitement du langage naturel. Lien : <https://aws.amazon.com/fr/what-is/nlp/>
- H.Belhadef, I.Djemal : Méthode TALN – Cours de l'unievrsite de Msila - Algérie
- B. Prieur : Traitement automatique du langage naturel avec Python. Le NLP avec SpaCy et NLTK.
- Baddari Kamel et Herzallah Abdelkarim (2015) : La recherche de l'information, livre – OPU -Algérie.