الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي و البحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
كلية العلوم الدقيقة والتطبيقية
Faculté des sciences exactes et appliquées

Rapport du projet PSD

Domaine : Mathématique & Informatique

Option : Informatique

PRESENTE PAR :

- Bahmed Aymene Abderrahmene
- Hamzaoui Fatima Ferdaousse
- Kamraoui Ismail
- Bordji Mohammed

THÈME :

L'étude de l'impact de l'utilisation des réseaux sociaux sur la santé mentale des étudiants

MASTER DATA SCIENCE

ANNEE UNIVERSITAIRE : 2025/2026

1. Introduction

Ce mini-projet vise à analyser l'impact de l'utilisation des réseaux sociaux sur la santé mentale des étudiants. Le dataset choisi porte sur 905 étudiants et comprend 13 variables décrivant les caractéristiques démographiques, les habitudes d'utilisation des réseaux sociaux, le sommeil, la santé mentale, les conflits sociaux et le niveau d'addiction. L'objectif global est d'explorer les relations entre ces variables, de réduire la dimension des données par Analyse en Composantes Principales (ACP) et d'identifier des profils d'étudiants à l'aide de méthodes de clustering.

2. Nettoyage & Exploration

Étapes de nettoyage :

Le dataset a été fourni déjà nettoyer dès le début du projet. L'analyse des fonctions df.info() et df.head() confirme l'absence de valeurs manquantes ainsi que la cohérence des types de données. Aucune transformation lourde n'a été nécessaire.

Variables retenues :

Dans ce travail, toutes les variables du dataset ont été conservées et utilisées pour l'analyse. Étant donné que le dataset était déjà propre, aucune variable n'a été supprimée. L'ensemble des variables a été exploité afin de garantir une analyse globale et cohérente.

Visualisations :

Afin d'explorer et de comprendre les données, plusieurs visualisations ont été réalisées :

- Histogrammes : pour analyser la distribution des variables ;
- Boxplots : pour étudier la dispersion et détecter d'éventuelles valeurs extrêmes ;
- Barplots : pour comparer certaines catégories ;
- Scatterplots : pour observer les relations entre paires de variables ;
- Matrice de corrélation : pour synthétiser les relations linéaires entre toutes les variables ;
- Carte des individus (ACP) : pour visualiser la projection des étudiants sur les axes principaux ;
- Cercle de corrélation (ACP) : pour interpréter la contribution des variables aux axes factoriels ;
- Dendrogramme (HC) : pour analyser la structure hiérarchique des données ;
- Carte des individus (ACP + K-means) : Pour visualiser la répartition des clusters de K-means.

Observations principales : Les graphiques montrent une utilisation relativement élevée des réseaux sociaux chez une grande partie des étudiants, associée à un nombre d'heures de sommeil

inférieur aux recommandations, ce qui suggère un possible impact négatif sur le mode de vie et la santé mentale.

3. Analyse Statistique & ACP

Résultats describe() :

L'analyse du tableau généré par df.describe() permet de synthétiser les principales caractéristiques statistiques des variables numériques du dataset. Les résultats indiquent que l'échantillon est majoritairement composé d'étudiants en âge universitaire, avec un âge moyen de 20 ans. Le temps moyen consacré aux réseaux sociaux est de 5,1 heures par jour, traduisant un usage important. La durée moyenne de sommeil est de 6,64 heures par nuit, inférieure aux recommandations usuelles. Par ailleurs, les scores moyens de santé mentale, de conflits liés aux réseaux sociaux et d'addiction révèlent respectivement un niveau émotionnel modéré, des tensions modérées et un degré d'addiction relativement élevé chez les étudiants.

Corrélations :

La corrélation positive entre le temps d'utilisation des réseaux sociaux et le score d'addiction signifie que lorsque l'utilisation augmente, l'addiction augmente également. À l'inverse, une corrélation négative est observée entre les heures de sommeil et le niveau d'addiction.

Analyse en Composantes Principales (ACP) :

Une ACP a été appliquée sur les variables standardisées afin de réduire la dimension des données. La carte des individus permet de visualiser la répartition globale des étudiants, tandis que le cercle de corrélation met en évidence les relations entre les variables et les axes factoriels.

L'axe 1 oppose principalement une forte utilisation des réseaux sociaux et un niveau élevé d'addiction à de meilleures habitudes de sommeil. L'axe 2 est essentiellement porté par la variable âge, indiquant que cet axe reflète principalement des différences démographiques entre les étudiants. Les deux premiers axes expliquent une part importante de la variance totale.

4. Clustering (HC & K-means)

Méthodes utilisées :

Un clustering K-means a été réalisé sur les données projetées par l'ACP pour regrouper ses résultats. Ensuite, un clustering hiérarchique (CH) a été appliqué au jeu de données original. Suite à cette analyse, un clustering K-means a cette fois été effectué sur les données brutes.

Choix du nombre de clusters :

Le dendrogramme (HC) et la méthode du coude (K-means) ont unanimement suggéré un nombre optimal de 3 clusters et 2 clusters pour ACP + K-means.

Interprétation des clusters :

(ACP + K-means) :

- ***Cluster 1*** : Étudiants avec un bon mode de vie.
- ***Cluster 2*** *:* Étudiants avec un mauvais mode de vie.

(HC + K-means) :

- ***Cluster 1*** : Étudiants avec une utilisation élevée des réseaux sociaux, un score d'addiction élevé et un sommeil réduit.
- ***Cluster 2*** : Étudiants modérés, avec une utilisation équilibrée et une santé mentale moyenne.
- ***Cluster 3*** : Étudiants avec une faible utilisation, un meilleur sommeil et un score d'addiction plus bas**.**

5. Conclusion

Résumé :

Ce mini-projet a permis d'analyser l'impact de l'utilisation des réseaux sociaux sur le mode de vie et la santé mentale des étudiants. Les analyses statistiques, l'ACP et le clustering ont mis en évidence des profils distincts d'étudiants en fonction de leurs comportements numériques.

Limites :

L'étude repose sur des données déclaratives, susceptibles d'introduire des biais. De plus, l'analyse est transversale et ne permet pas d'établir des relations de causalité.

Améliorations possibles :

L'utilisation de données recueillies sur plusieurs périodes, l'ajout de variables comme la performance académique ou le niveau de stress, ainsi que la comparaison avec d'autres méthodes de clustering permettraient d'améliorer l'analyse.