



Université de Tunis El Manar  
République Tunisienne  
Ministère de l'Enseignement  
Faculté des Sciences de Tunis  
Supérieur  
et de la Recherche Scientifique  
Licence en Mathématique appliquées  
Université de Tunis El Manar  
Faculté des Sciences de Tunis  
Spécialité : Sciences des données

**End of Year project II report**

---

## Project Title

---

*Elaboré par:*

Moataz Ben Trad

*Class:*

2<sup>nd</sup> year INFO/TEL

**Encadrement**

Encadrante académique: Dr.Samir Toumi

Encadrante professionnel : Mr.Mohamed Charaabi

---

Academic Year: 2023/2024

# Acknowledgment

We would like to pay our special regards ...

We would also like to thank ...

# Abstract

Write your English abstract here

**Keywords** : Natural language processing, Artificial intelligence, Keyword, Keyword

## Résumé

Résumé en français ...

**Mots clés :** Traitement Automatique du Langage Naturel, Intelligence Artificielle,  
Mot clé, Mot clé

# Contents

<b>List of Figures</b>	<b>6</b>
<b>List of Tables</b>	<b>7</b>
<b>Acronyms</b>	<b>8</b>
<b>General introduction</b>	<b>9</b>
<b>1 Contexte général du projet:</b>	<b>11</b>
Introduction . . . . .	11
1.1 Présentation de l'organisme d'accueil : . . . . .	11
1.1.1 AI BRIGHT . . . . .	11
1.1.2 Activités d'Ai Bright : . . . . .	12
1.2 Présentation du projet . . . . .	13
1.2.1 description du projet . . . . .	13
1.2.2 Énoncé du problème . . . . .	13
1.2.3 Etude de l'existant: . . . . .	13
1.2.4 ritique de l'Existence et Avantages de Notre solution . . . . .	15
1.3 choix des méthodologies . . . . .	15
1.3.1 Étude comparative des méthodologies existantes . . . . .	15
1.3.2 Présentation de la méthodologie retenue . . . . .	16
Conclusion . . . . .	17
<b>2 COMPREHENSION METIER:</b>	<b>18</b>
Introduction . . . . .	18
2.1 Objectif métiers . . . . .	18
2.2 Objectif sciences de données . . . . .	19
2.3 Objectif sciences de données . . . . .	19
2.3.1 Matériel . . . . .	19
2.3.2 Logicies . . . . .	19
Conclusion . . . . .	21

<b>3</b>	<b>Compréhension et préparation des données</b>	<b>22</b>
	Introduction . . . . .	22
3.1	Compréhension des données . . . . .	22
3.1.1	Sources de Données . . . . .	22
3.1.2	Objectifs de Compréhension des Données . . . . .	23
3.2	Section . . . . .	23
3.2.1	Subsection title . . . . .	23
3.2.2	Subsection title . . . . .	24
	Conclusion . . . . .	24
<b>4</b>	<b>Implementations and achievements</b>	<b>25</b>
	Introduction . . . . .	25
4.1	XX Dataset . . . . .	25
4.1.1	Data exploration . . . . .	25
4.1.2	Data cleaning . . . . .	25
4.2	Experimental Results . . . . .	25
4.2.1	Results of XX Dataset . . . . .	26
4.2.2	Results of YY Dataset . . . . .	26
4.2.3	Comparison between XX and YY datasets . . . . .	26
	Conclusion . . . . .	26
	<b>Conclusion</b>	<b>27</b>
	<b>Bibliography</b>	<b>28</b>

# List of Figures

1.1	AI Bright Logo . . . . .	11
1.2	AIMOOV . . . . .	12
1.3	Fitbit . . . . .	14
1.4	Apple Health . . . . .	14
1.5	Google Fit . . . . .	15
1.6	Méthodologie CRISP-DM . . . . .	16
2.1	Python Logo . . . . .	19
2.2	Anaconda Logo . . . . .	20
2.3	Pytorch Logo . . . . .	20
2.4	Google colab Logo . . . . .	20
2.5	Git Logo . . . . .	21
2.6	Discord Logo . . . . .	21



# List of Tables

1.1	The training time (in minutes) on 1 and 8 workers with the corresponding speedup . . . . .	17
-----	--	----

# List of Acronyms

- AI:

- :

- :

- :

- :

- :

- :

- :

- :

- :

- :.

- :

- :

# Introduction générale :

Le domaine de la e-santé et du e-fitness a grandement bénéficié de l'évolution des technologies numériques. L'intégration rapide de l'intelligence artificielle dans ce secteur a permis de développer des applications de fitness à domicile qui sont devenues cruciales, notamment en période de pandémie. Le projet AIMOOV d'AI BRIGHT vise à exploiter ces technologies pour offrir une expérience de fitness sur mesure. Ce projet s'inscrit dans le contexte de l'intégration progressive des données de smartwatches et smartphones, visant à améliorer la précision et l'efficacité des entraînements à domicile..

La nouvelle version de notre produit est conçue pour utiliser les données des smartwatches et des écosystèmes comme Google Fit et Apple Health. Cependant, nous faisons face à un défi important : l'absence de données historiques suffisantes pour entraîner nos modèles d'intelligence artificielle. Pour résoudre ce problème, nous avons mis en place des méthodes de génération de données. Cette solution nous permettra d'utiliser les données réelles dès qu'elles seront disponibles, sans attendre une longue période de collecte. Notre but est de créer une expérience de fitness très personnalisée et efficace, qui s'adapte dynamiquement aux besoins changeants de l'utilisateur, en utilisant des techniques d'apprentissage automatique..

Nous adoptons une approche hybride combinant la simulation de données et l'apprentissage machine pour préparer le système à l'utilisation future des données réelles. Cette méthodologie inclut le développement de modèles prédictifs et adaptatifs qui peuvent être mis à jour dynamiquement avec de nouvelles données.

Le projet couvrira le développement du modèle initial, la création d'un simulateur de données, et l'intégration de ces éléments dans une plateforme fonctionnelle. Les livrables incluront un prototype de système, des rapports de validation de modèle, et une documentation complète sur les processus et les méthodes utilisés.

Ce rapport synthétise l'ensemble des travaux réalisés dans cette perspective. Il est organisé en cinq chapitres comme suit :

---

# Chapter 1

## Contexte général du projet:

### Introduction

Tout d’abord, nous commençons par présenter l’entreprise d’accueil et ses activités. Ensuite, nous poursuivons avec le contexte du projet : sa description, l’énoncé du problème en annonçant notre solution et nos objectifs, ainsi que les travaux connexes. Enfin, nous présentons la méthodologie de gestion du projet.

### 1.1 Présentation de l’organisme d’accueil :

#### 1.1.1 AI BRIGHT

Fondée en 2019 et basée à Levallois-Perret, Paris, la startup AI Bright innove dans le secteur sportif en exploitant le potentiel de l’intelligence artificielle. Son objectif est de rendre l’IA accessible à tous grâce à des applications innovantes conçues pour améliorer la santé et le bien-être quotidien.



Figure 1.1: AI Bright Logo

---

AI Bright développe des solutions personnalisées qui utilisent des techniques avancées en apprentissage automatique pour rendre la pratique sportive plus intuitive et adaptée aux besoins individuels de ses utilisateurs. En combinant l'IA et la réalité Augmentée, AI Bright propose une interface interactive qui simplifie et optimise l'entraînement.. L'ambition d'AI Bright est de devenir un leader mondial dans le domaine la vidéo intelligente, en ciblant à la fois les marchés grand public et professionnel

### 1.1.2 Activités d'Ai Bright :

AI BRIGHT a développe « AIMOOV », une caméra intelligente qui offre les avantages d'un entraîneur à domicile, avec une différence notable : votre coach est l'Intelligence Artificielle (IA) .AIMOOV est conçu pour être le plus petit instructeur de gym du monde. Cette caméra intelligente vous aide à vous entraîner chez vous en suivant vos progrès et en vous corrigeant lorsque vous vous entraînez mal. Elle fonctionne en temps réel et se synchronise avec vos applications de suivi de santé



Figure 1.2: AIMOOV

AI BRIGHT intègre plusieurs technologies avancées pour enrichir et optimiser cette caméra intelligente : Vision par ordinateur permet à la caméra de classifier et de régresser des images et vidéos, offrant ainsi des fonctionnalités avancées en temps réel Reconnaissance vocale : Cette technologie permet aux utilisateurs d'interagir facilement avec AIMOOV, facilitant son utilisation par une grande variété de publics. Réalité augmentée enrichit l'environnement immédiat de l'utilisateur avec des objets virtuels intégrés de manière réaliste, ce qui améliore l'engagement. Animation 3D : Le logiciel d'AIMOOV peut transformer n'importe quelle vidéo en une animation

---

3D, grâce à l'apprentissage profond, permettant des analyses et des visualisations dynamiques. Ces technologies ne se limitent pas à enrichir l'expérience utilisateur ; elles transforment AIMOOV en un outil robuste et polyvalent pour le fitness à domicile,

## **1.2 Présentation du projet**

### **1.2.1 description du projet**

En décembre, Ai Bright a lancé AIMOOV, une caméra intelligente conçue pour personnaliser et optimiser l'expérience de fitness à domicile. Ce dispositif intégrera de façon sécurisée des données provenant de Google Fit et Apple Health, y compris des indicateurs physiologiques et d'activité tels que le rythme cardiaque, le nombre de pas et les activités externes. Ces données seront collectées en temps réel via des dispositifs connectés, tels que les smartphones et les montres intelligentes, avec le consentement de l'utilisateur. Dans ce contexte, notre objectif sera de fournir une expérience de fitness hautement personnalisée et efficace, adaptant dynamiquement chaque session aux besoins évolutifs de l'utilisateur, en utilisant des techniques d'apprentissage automatique.

### **1.2.2 Énoncé du problème**

La nouvelle version de notre produit, conçue pour intégrer les données de smartwatches et des écosystèmes comme Google Fit et Apple Health, est confrontée à un défi majeur : le manque de données historiques pour entraîner efficacement nos modèles d'intelligence artificielle. Ce déficit rend difficile la création immédiate d'une expérience de fitness personnalisée et efficace. Notre objectif est de développer des solutions qui permettent d'utiliser des techniques d'apprentissage automatique pour adapter dynamiquement chaque session de fitness aux besoins évolutifs de chaque utilisateur, sans avoir à attendre une longue période de collecte de données initiale.

### **1.2.3 Etude de l'existant:**

L'étude de l'existant est une étape préliminaire avant de se lancer dans tout projet. Elle nous donne un aperçu de la pertinence du projet, de sa faisabilité et de sa durabilité.

#### **Fitbit**

Fitbit propose des montres intelligentes et des bracelets de fitness qui collectent des

---

données telles que le rythme cardiaque, les pas, la qualité du sommeil, et les calories brûlées. L'IA est utilisée pour analyser ces données et fournir des recommandations personnalisées pour améliorer la santé et la forme physique des utilisateurs.



Figure 1.3: Fitbit

### **Apple**

Avec l'Apple Watch, Apple offre une expérience de fitness intégrée qui inclut le suivi des activités, le coaching de santé, et la surveillance médicale. L'Apple Watch utilise l'IA pour interpréter les données de santé et d'activité, offrant des rappels personnalisés pour bouger ou s'entraîner.

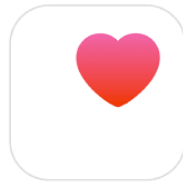


Figure 1.4: Apple Health

### **Google**

Google, à travers ses Wear OS et ses intégrations avec Google Fit, permet aux utilisateurs de suivre divers indicateurs de santé et de fitness. Google Fit utilise des algorithmes d'IA pour fournir des insights basés sur les données collectées, comme les minutes d'activité et les points cardio. Ces fonctionnalités aident les utilisateurs à comprendre leur niveau d'activité physique et à recevoir des conseils personnalisés pour améliorer leur condition physique et leur bien-être.





Figure 1.5: Google Fit

#### **1.2.4 Critique de l'Existence et Avantages de Notre solution**

Bien que les produits de Fitbit, Apple et Google offrent des fonctionnalités avancées pour le suivi de la santé et du fitness, malgré la collecte d'une grande variété de données, leur utilisation pour personnaliser les entraînements en temps réel reste généralement superficielle. Les programmes proposés sont souvent standards et ne s'adaptent pas dynamiquement aux performances et aux progrès individuels de l'utilisateur au cours d'une session. En contraste, notre produit se distingue par son accès direct à ces données, ainsi que par sa capacité à offrir des entraînements personnalisés selon l'objectif et la condition physique de l'utilisateur. Grâce à l'intégration de notre caméra intelligente, non seulement nous suivons les données physiologiques de l'utilisateur en temps réel, mais nous corrigeons également les mouvements de l'utilisateur pendant ses exercices. Cette fonctionnalité de correction en temps réel permet une adaptation et une amélioration immédiates, guidant l'utilisateur de manière efficace pour atteindre ses objectifs de fitness de manière plus précise et engagée.

### **1.3 choix des méthodologies**

Il existe plusieurs méthodologies de gestion qui peuvent être appliquées aux projets de Data Science. Les plus courantes sont CRISP-DM, SEMMA, KDD. Dans cet article, nous allons les comparer afin de choisir la méthodologie la plus adéquate.

#### **1.3.1 Étude comparative des méthodologies existantes**

Nous présentons dans cette sous-section une comparaison entre les méthodologies CRISPDM, SEMMA et KDD.

---

### 1.3.2 Présentation de la méthodologie retenue

Après avoir comparé les différentes méthodologies, nous avons finalement opté pour la méthodologie CRISP-DM. Cette méthode, agile et itérative, est la plus appropriée pour notre projet car elle correspond mieux à nos besoins en termes de construction, de déroulement des opérations. La méthode CRISP, connue sous le nom de CRISP-DM, a été développée par IBM dans les années 1960 pour les projets de data mining. Aujourd'hui, c'est l'une des méthodes les plus performantes pour les projets de science des données. Elle se décompose en 6 étapes allant de la compréhension du problème métier au déploiement et la mise en production :

**1-Compréhension du problème métier :**

Cette étape a pour objectif d'obtenir une vue d'ensemble de l'activité du client.

**2- Compréhension des données:**

déterminez les données pour répondre à vos besoins.

**3- Prétraitement des données :**

préparer les données dans un format adéquat.

**4- Modélisation :**

concevoir des modèles de données/solutions pour répondre à vos besoins.

**5- Evaluation :**

tester les résultats par rapport aux objectifs du projet.

**6- Déploiement :**

déployer la solution et/ou les résultats dans une application

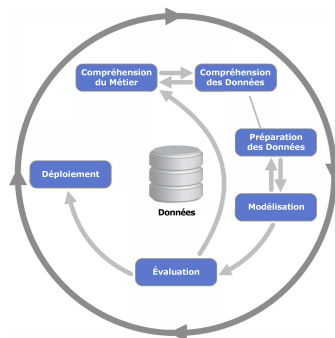


Figure 1.6: Méthodologie CRISP-DM

## Conclusion

Après avoir présenté le projet et l'organisation hôte, nous suivrons dans les chapitres suivants les étapes de la méthodologie "CRISP-DM" adoptée, de la

---

Epoch

Table 1.1: The training time (in minutes) on 1 and 8 workers with the corresponding speedup

compréhension métier au déploiement.

# Chapter 2

## COMPREHENSION METIER:

### Introduction

La première étape de notre méthodologie consiste à comprendre ce que nous voulons réaliser d'un point de vue opérationnel ainsi que les objectifs en matière de science des données.

### 2.1 Objectif métiers

Dans cette section, nous définissons les objectifs métiers de notre projet, qui sont cruciaux pour assurer une compréhension approfondie des exigences commerciales et de l'impact potentiel de notre solution. AI BRIGHT vise à révolutionner le fitness à domicile en lançant une solution de fitness intelligente et interactive. L'objectif de ce projet est de fournir une expérience utilisateur enrichie et personnalisée grâce à l'intégration intelligente des données de santé provenant de divers dispositifs connectés.

Le succès de ce projet sera mesuré par les critères suivants :

Lancement réussi sur le marché : Mise en place et adoption d'une solution de fitness à domicile qui utilise l'intelligence artificielle pour offrir une personnalisation sans précédent. Satisfaction des utilisateurs : Amélioration notable de l'expérience utilisateur, avec des retours positifs quant à l'efficacité et la personnalisation des sessions d'entraînement. Intégration des données de santé : Capacité à synchroniser de manière sécurisée et efficace les données provenant des écosystèmes tels que Google Fit et Apple Health, augmentant ainsi la valeur ajoutée pour l'utilisateur.

---

## 2.2 Objectif sciences de données

Afin d'atteindre les objectifs business cités ci-dessus, Simulation de données : Développer des méthodes pour générer ou simuler des données qui compensent le manque de données historiques. Cela permettra une mise en service rapide des modèles d'IA sans attendre la collecte prolongée de données, garantissant ainsi une amélioration continue et une personnalisation basée sur des scénarios réalistes. Optimisation des objectifs de fitness : Utiliser les données pour aider les utilisateurs à atteindre leurs objectifs de fitness de manière plus efficace. Le modèle analysera les tendances des données d'entraînement et des dispositifs connectés pour proposer des ajustements dynamiques qui maximisent les résultats des utilisateurs en fonction de leurs objectifs spécifiques.

## 2.3 Objectif sciences de données

### 2.3.1 Matériel

Pour notre projet, nous utilisons un Victus doté d'une mémoire RAM de 32 GO. Le processeur utilisé est un AMD Ryzen™ 5-5600H avec une carte graphique NVIDIA® GeForce® GTX 1650 4 GB GDDR6 et un disque dur SSD M.2 TLC NVMe™ PCIe® Gen 3.0 x4 de 512 GB.

### 2.3.2 Logiciels

Notre travail est réalisé sous le système d'exploitation Windows 11. Pour notre projet, nous utilisons :

Python : la partie modélisation est écrite en Python, dont le logo est présenté à la figure 2.1. C'est un langage de programmation interprété, orienté objet et de haut niveau. Il vise à aider les programmeurs à écrire un code clair et logique pour des projets de petite et grande envergure.



Figure 2.1: Python Logo

Anaconda : une distribution open source qui est la manière la plus facile et la plus efficace de faire de la science des données avec Python.



Figure 2.2: Anaconda Logo

PyTorch est une bibliothèque open-source de machine learning utilisée pour développer et former des modèles de réseaux de neurones. Elle offre des outils flexibles pour la recherche en intelligence artificielle et le déploiement de solutions d'apprentissage automatique.



Figure 2.3: Pytorch Logo

Google Colab: Google Colab ou Colaboratory est un service cloud, proposé par Google principalement en accès libre, basé sur Jupyter Notebook et destiné à la formation et à la recherche dans le processus d'apprentissage. Cette plateforme permet de former des modèles d'apprentissage automatique directement dans le cloud, sans avoir besoin d'installer des bibliothèques nécessaires ou autres sur notre ordinateur,



Figure 2.4: Google colab Logo

Git : C'est un système de contrôle de version distribué gratuit et open source conçu pour gérer des projets allant de petits à très grands avec rapidité et efficacité.

---

Git permet et encourage à avoir plusieurs branches locales qui peuvent être totalement indépendantes les unes des autres. La création, la fusion et la suppression de ces lignes de développement se font en quelques secondes.



Figure 2.5: Git Logo

Discord : Pour la communication collaborative, nous utilisons Discord, une plateforme qui permet des échanges textuels, vocaux et vidéo en temps réel. Elle nous aide à maintenir une communication fluide et régulière entre les membres de l'équipe, essentielle pour la coordination de nos activités de développement et pour les réunions d'avancement du projet. Discord offre également des fonctionnalités de partage d'écran et de groupes privés, ce qui renforce notre efficacité collaborative.



Figure 2.6: Discord Logo

## Conclusion

Conclusion Ce chapitre a été consacré à la compréhension métier, qui inclut les objectifs métier et les objectifs de data mining que nous cherchons à atteindre.

# Chapter 3

## Compréhension et préparation des données

### Introduction

Le présent chapitre est consacré aux phases de compréhension et de préparation des données de notre méthodologie CRISP-DM. Nous commençons par comprendre les données, ce qui est d'une importance vitale car cela permet d'éviter des problèmes inattendus lors de la phase suivante, la préparation des données. Ensuite, nous passons à la préparation des données qui comprendra plusieurs étapes différentes afin de d'effectuer toutes les transformations nécessaires pour obtenir notre ensemble de données final qui sera utilisé, ensuite, dans la phase de modélisation.

### 3.1 Compréhension des données

Dans cette section, nous allons détailler notre approche pour comprendre les données nécessaires à notre projet. Étant donné que nous n'avons pas accès aux vraies données pour le moment, nous utiliserons des méthodes de simulation pour créer un ensemble de données réalistes. Cette démarche est essentielle pour entraîner notre modèle d'intelligence artificielle et garantir que notre solution puisse s'adapter efficacement une fois les données réelles disponibles.

#### 3.1.1 Sources de Données

##### Données d'Entraînement

Description :

Les données d'entraînement incluent les informations personnelles et les détails spécifiques de chaque session d'exercice, enregistrées en temps réel grâce à la caméra



---

intelligente. Elles comprennent :

Identifiant : Un identifiant unique pour chaque utilisateur.

Informations démographiques : Âge Sexe Poids Taille Détails de l'exercice : Temps d'exercice pour chaque type d'exercice. Nombre de répétitions pour chaque exercice. Nombre d'erreurs majeures et mineures par séance Objectifs de l'entraînement : Type d'objectif de l'utilisateur (doux, force, just moov, perte de poids).

### **Données des Smartwatches**

Description : Les données collectées via les smartwatches fournissent des insights physiologiques et sur l'activité de l'utilisateur :

Rythme cardiaque

Nombre de pas

Calories brûlées

Niveau d'oxygène dans le sang (SpO2)

Température du corps

Analyse du sommeil

Remarque : Les mesures et la variété des données peuvent varier selon la marque de la smartwatch et nécessitent l'autorisation explicite de l'utilisateur pour être collectées et utilisées.

### **Données des Smartphones**

Description :

Les smartphones, via les écosystèmes tels que Apple Health et Google Fit, fournissent des données complémentaires qui enrichissent le profil de santé de l'utilisateur :

Données agrégées de santé et d'activité collectées par les applications de santé.

Remarque : L'accès à ces données nécessite également l'autorisation de l'utilisateur, garantissant ainsi la conformité avec les normes de confidentialité et de protection des données.

## **3.1.2 Objectifs de Compréhension des Données**

After choosing the ....

## **3.2 Section**

A brief introduction to the following subsections (**at least two sentences**).

### **3.2.1 Subsection title**

Our first step was to choose ....

---

### **3.2.2 Subsection title**

After choosing the ....

## **Conclusion**

In this chapter, We have gone over the environment we are working with as well as the results of .....

# Chapter 4

## Implementations and achievements

### Introduction

In this chapter, we will compare the XX dataset results with a different dataset, namely, YY which contains Tunisian and DD Dialects only. Then, we will compare ...

### 4.1 XX Dataset

A brief introduction to the following subsections (**at least two sentences**).

#### 4.1.1 Data exploration

#### 4.1.2 Data cleaning

### 4.2 Experimental Results

A brief introduction to the following subsections (**at least two sentences**).

---

#### **4.2.1 Results of XX Dataset**

#### **4.2.2 Results of YY Dataset**

#### **4.2.3 Comparison between XX and YY datasets**

##### **Discussion**

As a conclusion of the datasets comparison, ...

### **Conclusion**

In this chapter, we introduced the collected XX dataset as well as the other datasets used for the comparison. We exposed their results and we compared them to the results presented in the third chapter of the dataset provided in the CC competition.

# Conclusion

**Writing a general conclusion involves summarizing the main points of your work, reflecting on its significance, and offering any final thoughts or recommendations. Here's a structured approach to writing a general conclusion:**

1. Begin by summarizing the key findings or results of your work. Highlight the most important discoveries, insights, or conclusions that you have reached throughout your project or study.
2. Remind readers of the objectives or goals you set out to achieve at the beginning of your work. Discuss how well you have met these objectives and whether you have successfully addressed the problems you identified.
3. Acknowledge any limitations or constraints of your study. Discuss any challenges you encountered, such as methodological limitations, data constraints, or unexpected obstacles, and how these may have affected your results or conclusions.
4. Identify areas for future research or further investigation based on the findings of your work. Suggest potential approaches that could improve your work and open further perspectives.

# Bibliography