

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Ecole Nationale Polytechnique



Département de Génie Industriel

Mémoire de projet de fin d'études  
Pour l'obtention du diplôme d'ingénieur d'état en Management Industriel

**Développement d'un outil d'aide à la décision basé sur  
l'apprentissage non-supervisé pour l'élaboration d'une feuille de  
route  
- Application : Groupe SONATRACH**

Nazim TAKHEDMIT

Sous la direction de M. Mabrouk AIB (Dr ENP)

Présenté et soutenu publiquement le 21/06/2018

**Composition du Jury :**

Président	Mme Nacéra ABOUN	MAA	ENP
Promoteur	M. Mabrouk AIB	Dr	ENP
Examineur	Mme Sabiha NAIT KACI	MAA	ENP



REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Ecole Nationale Polytechnique



Département de Génie Industriel

Mémoire de projet de fin d'études  
Pour l'obtention du diplôme d'ingénieur d'état en Management industriel

**Développement d'un outil d'aide à la décision basé sur  
l'apprentissage non-supervisé pour l'élaboration d'une feuille de  
route  
- Application : Groupe SONATRACH**

Nazim TAKHEDMIT

Sous la direction de M. Mabrouk AIB (Dr ENP)

Présenté et soutenu publiquement le 21/06/2018

**Composition du Jury :**

Président	Mme Nacéra ABOUN	MAA	ENP
Promoteur	M. Mabrouk AIB	Dr	ENP
Examineur	Mme Sabiha NAIT KACI	MAA	ENP

## *Dédicaces*

*Je dédie ce travail à mes très chers  
parents, à mes deux frères, à mes amis  
et à toute personne qui a contribué  
directement ou indirectement à ce que je  
suis aujourd'hui.*

# *Remerciements*

Mes remerciements s'adressent à mon encadrant en entreprise et mon promoteur Monsieur Aib Mabrouk, pour ses précieux conseils, son aide ainsi que pour toutes ses qualités.

Je remercie ensuite toute l'équipe SAP, avec qui j'ai eu le plaisir d'apprendre et de travailler pendant mon stage.

Je tiens également à remercier tous les enseignants du Département Génie Industriel de l'Ecole Nationale Polytechnique de m'avoir assuré la formation d'ingénieur, et tous les rudiments qui ont rendu ce travail possible.

Enfin je remercie tous les membres du jury qui m'ont fait l'honneur d'évaluer mon travail

### ملخص:

الهدف من هذا العمل هو تطوير برنامج آلي باستعمال تقنية من تقنيات الذكاء الاصطناعي التي هي التعلم الغير مراقب، وهذا إعتقادا على خوارزمية k-means، و هذا لغرض المساهمة في استراتيجية مجمع سوناتراك المتعلقة بمخطط عمل حول تحديث انظمة معلوماتها عبر برامج SAP التي قامت بشرائها حصريا.

الكلمات المفتاحية: k-means، مخطط عمل، تجمع، مركز ثقل، اتخاذ القرار، برامج SAP، تصنيف.

### ABSTRACT:

The aim of this work is to develop a decision-making tool, based on unsupervised learning, which is an artificial intelligence technique, by using k-means algorithm. This tool will analyze all the gathered data regarding SONATRACH subsidiaries and their information systems, in order to provide a roadmap that will assist the group in its strategy of implementing SAP software solutions.

**Key Words:** clusters, k-means, roadmap, centroids, decision-making, SAP software solutions, classification.

### Résumé :

Ce travail consiste à développer un outil d'aide à la décision basé sur une technique d'intelligence artificielle, qui est l'apprentissage non-supervisé, cela en utilisant l'algorithme des k-moyennes. Ensuite, cet outil va servir comme moyen, pour analyser un ensemble de données relatives aux filiales de SONATRACH, ainsi qu'à leur système d'information, dans le but d'établir une feuille de route, qui va orienter le groupe SONATRACH dans sa mise en place de solutions SAP récemment acquises.

**Mots clés :** clustering, k-moyennes, feuille de route, centroïdes, aide à la décision, solutions SAP, classification.

## **Table des matières**

### **Liste des figures**

### **Liste des tableaux**

### **Liste des abréviations**

<b>Introduction générale.....</b>	<b>10</b>
<b>Chapitre 1 : Etat des lieux</b>	
<b>Introduction .....</b>	<b>12</b>
<b>1.1 SAP SE .....</b>	<b>13</b>
<b>1.1.1 présentation de l'entreprise.....</b>	<b>13</b>
<b>1.1.1.1 historique.....</b>	<b>13</b>
<b>1.1.1.2 SAP Aujourd'hui.....</b>	<b>13</b>
<b>1.1.2 industries .....</b>	<b>14</b>
<b>1.1.3 le portefeuille de solutions .....</b>	<b>14</b>
<b>1.2 SONATRACH .....</b>	<b>16</b>
<b>1.2.1 présentation du groupe .....</b>	<b>16</b>
<b>1.2.2 filiales concernées par l'étude .....</b>	<b>19</b>
<b>1.2.2.1 NAFTAL .....</b>	<b>19</b>
<b>1.2.2.2 ENAFOR .....</b>	<b>20</b>
<b>1.2.2.3 HYPROC.....</b>	<b>21</b>
<b>1.3 Le contexte de l'étude et de la problématique .....</b>	<b>21</b>
<b>1.4 Conclusion.....</b>	<b>22</b>
<b>Chapitre 2 : Etat de l'art</b>	
<b>Introduction .....</b>	<b>23</b>
<b>2.1 Système d'informations .....</b>	<b>24</b>
<b>2.1.1 définitions.....</b>	<b>24</b>
<b>2.1.2 types des SI.....</b>	<b>25</b>
<b>2.1.2 types de solutions informatiques pour la gestion des SI .....</b>	<b>25</b>
<b>2.2 Intelligence artificielle.....</b>	<b>28</b>
<b>2.2.1 définitions.....</b>	<b>28</b>
<b>2.2.2 les sous-domaines de l'IA.....</b>	<b>29</b>
<b>2.3 Machine learning (apprentissage machine) .....</b>	<b>29</b>
<b>2.3.1 définitions.....</b>	<b>29</b>
<b>2.3.2 les types d'apprentissages .....</b>	<b>30</b>
<b>2.4 Clustering.....</b>	<b>30</b>
<b>2.4.1 définitions.....</b>	<b>30</b>
<b>2.4.2 la norme de Minkowski.....</b>	<b>31</b>
<b>2.4.3 les algorithmes du clustering.....</b>	<b>31</b>
<b>2.5 Aide multicritère à la décision.....</b>	<b>34</b>

2.5.1 définitions.....	34
2.5.2 les types de problématiques.....	35
2.5.3 l'aide à la décision et l'apprentissage .....	35
2.6 Conclusion.....	35
<b>Chapitre 3 Développement de la solution</b>	
Introduction .....	36
3.1 Les étapes du développement de la solution .....	37
3.1.1 développement du logiciel de clustering.....	37
3.1.1.1 la définition des besoins .....	37
3.1.1.2 la modélisation de l'architecture.....	37
3.1.1.3 codage et programmation.....	40
3.1.1.4 la validation.....	40
3.1.2 la recherche des données.....	41
3.1.2.1 les données sur les filiales.....	41
3.1.2.2 les données sur les SI des filiales .....	43
3.1.3 l'analyse des résultats.....	44
3.1.3.1 l'analyse du clustering des données sur les filiales.....	44
3.1.3.2 l'analyse du clustering des données sur les SI des filiales.....	46
3.2 L'élaboration de la feuille de route.....	51
3.3 Conclusion.....	52
Conclusion générale .....	53
Bibliographie.....	54
Annexes .....	56

## Liste des figures :

<b>figure 1: SAP aujourd'hui .....</b>	<b>13</b>
<b>figure 2: portefeuille de solutions SAP (résumé) .....</b>	<b>14</b>
<b>figure 3: la plateforme cloud de SAP.....</b>	<b>15</b>
<b>figure 4: organigramme du groupe SONATRACH.....</b>	<b>17</b>
<b>figure 5: les filiales du groupe SONATRACH par holdings .....</b>	<b>18</b>
<b>figure 6: le passage d'une donnée à une information.....</b>	<b>24</b>
<b>figure 7: les couches décisionnelles au sein d'une entreprise .....</b>	<b>25</b>
<b>figure 8: les différentes solutions informatiques d'une entreprise en fonction des couches décisionnelles.....</b>	<b>27</b>
<b>figure 9: le processus de la BI.....</b>	<b>28</b>
<b>figure 10: l'évolution des clusters et de leurs centroïdes dans l'algorithme des k-moyennes .....</b>	<b>34</b>
<b>figure 11: l'architecture du programme informatique développé.....</b>	<b>37</b>
<b>figure 12: les entrées.....</b>	<b>38</b>
<b>figure 13: le traitement .....</b>	<b>39</b>
<b>figure 14: les résultats ou sorties.....</b>	<b>39</b>
<b>figure 15: la couche « utilisateur ».....</b>	<b>40</b>
<b>figure 16: interface graphique principale .....</b>	<b>41</b>
<b>figure 17: résultat du clustering des filiales .....</b>	<b>45</b>
<b>figure 18: rapport sur le clustering des filiales .....</b>	<b>45</b>
<b>figure 19: résultat du clustering pour NAFTAL .....</b>	<b>46</b>
<b>figure 20: rapport du clustering pour NAFTAL .....</b>	<b>47</b>
<b>figure 21: résultat du clustering pour ENAFOR.....</b>	<b>48</b>
<b>figure 22: rapport du clustering d'ENAFOR .....</b>	<b>48</b>
<b>figure 23: résultat du clustering pour HYPROC .....</b>	<b>49</b>
<b>figure 24: rapport du clustering d'HYPROC .....</b>	<b>50</b>
<b>figure 25: schéma explicatif de la feuille de route .....</b>	<b>52</b>

## Liste des Tableaux :

Tableau 1: données des différentes filiales (CA, résultat, nombre d'employés, année de création) .....	42
Tableau 2: données sur la maturité des SI de NAFTAL, ENAFOR et HYPROC.....	43

## **Abréviations :**

**SI:** Système d'Information

**IA:** Intelligence Artificielle

**CRM:** Customer Relationship Management

**BI:** Business Intelligence

**APS:** Advanced Planning System

**ERP:** Enterprise Resource Planning

**SRM:** Supplier Relationship Management

**SCEM:** Supply Chain Event Management

**MES:** Manufacturing Execution System

**CSV:** Comma Separated Values

**ETL:** Extract Transform Load

# Introduction générale

---

L'ère de la digitalisation avancée que vit le monde d'aujourd'hui, a créé un écosystème dans lequel toute entreprise qui veut survivre doit s'aligner avec ses exigences. Toutefois, cela fait l'objet d'une stratégie qu'elle doit mettre en place, le plus tôt possible, pour ne pas être victime par la suite d'obsolescence ou d'une grande résistance au changement, qui va rendre ce type de projet irréalisable ou entraîné dans un échec certain.

Le groupe SONATRACH a bien compris les contraintes dudit écosystème, et a entamé l'étude d'une stratégie, par laquelle le groupe vivra une restructuration totale, avec de nouvelles réformes sur tous les plans. Cette tâche sera assurée en collaboration avec des bureaux de conseil et d'autres partenaires internationaux, spécialisés dans chaque type de réforme. Parmi ces dernières, la volonté d'unifier les workflows dans un seul système centralisé et, à cet effet, SAP a été choisie pour assurer ce travail.

SAP considère l'Afrique comme un marché stratégique, où beaucoup d'opportunités de développement existent. Cela étant dû au nombre important d'entreprises dont les systèmes d'information sont peu matures, et qui ont besoin d'un accompagnement et de l'acquisition d'ERP et de solutions de nouvelle génération, afin de contribuer au développement de l'entreprise elle-même d'une part, et du continent d'autre part.

Le groupe SONATRACH fait partie de ces entreprises, et compte tenu du fait qu'il soit parmi les plus grandes entreprises en Afrique, SAP le considère comme un partenaire stratégique par excellence pour bénéficier d'une transformation digitale.

Le travail effectué, s'inscrit dans le cadre d'une stratégie de mise en place de solutions du portefeuille SAP, récemment acquis par le groupe. Cette étude fera intervenir des outils de différentes disciplines, mais dans un but unique, qui est l'obtention d'une stratégie et d'un plan d'action efficaces.

Ledit travail est structuré selon des chapitres, qui sont les suivants :

## Chapitre 1 : Etat des lieux

Dans ce chapitre, il y aura la présentation des différentes entreprises, avec leurs activités et leur structuration, ainsi que leurs différents enjeux. Il y sera également défini le contexte de l'étude.

## Chapitre 2 : Etat de l'art

Ce chapitre va inclure toutes les bases académiques et les présentations des différents outils et techniques, qui seront utilisés par la suite pour mettre en pratique le travail envisagé.

## Chapitre 3 : Développement de la solution

Cette partie s'intéresse à la solution proposée. Au début, elle montre les différentes étapes du développement de l'outil d'aide à la décision, viennent après la recherche et la présentation des données, suivies par l'utilisation de l'outil d'aide à la décision, et l'analyse des résultats de ce dernier. Pour finir, une recommandation sera exprimée par un plan d'action.

Cette partie sera suivie d'une synthèse en guise de conclusion générale. Cette dernière va reprendre l'essentiel du travail ainsi que les éventuelles pistes d'améliorations.

# Chapitre 1

## Etat des lieux

---

### Introduction

Ce chapitre décrit les deux entreprises, qui sont concernées par l'étude, il vise aussi la présentation de leurs challenges et ambitions. Le besoin de la transformation digitale et aussi celui de l'alignement sur les bonnes pratiques internationales, a conduit le groupe SONATRACH à revoir ses solutions informatiques dédiées à ses systèmes d'information. Pour cela, elle a consulté SAP, qui propose un portefeuille de solutions très riche, et qui dispose d'une expertise dans le domaine des systèmes d'information, nécessaire pour accompagner le groupe dans cet axe stratégique.

## 1.1 SAP SE

### 1.1.1 présentation de l'entreprise

SAP (Systems, Applications and Products for data processing en anglais et Systeme, Anwendungen und Produkte in der Datenverarbeitung en allemand):

SAP est une entreprise de droits européens, dont le siège social est basé à Waldorf en Allemagne, elle offre des services et des solutions informatiques pour entreprises.

#### 1.1.1.1 historique

L'entreprise SAP a été fondée en 1972 en Allemagne, par cinq anciens ingénieurs d'IBM. Ils avaient pour vision, l'exploitation du potentiel immense de l'informatique dans le domaine des affaires. Ils avaient conçu un logiciel capable de gérer les données, à destination des entreprises.

#### 1.1.1.2 SAP Aujourd'hui

SAP aujourd'hui, est une multinationale qui compte 183 000 clients, et parmi ces derniers, des entreprises de renom. Elle offre une gamme de produits et de services assez variés. Elle a développé une expertise dans différentes industries, ce qui l'a rendue leader dans le domaine. (voir figure 1)

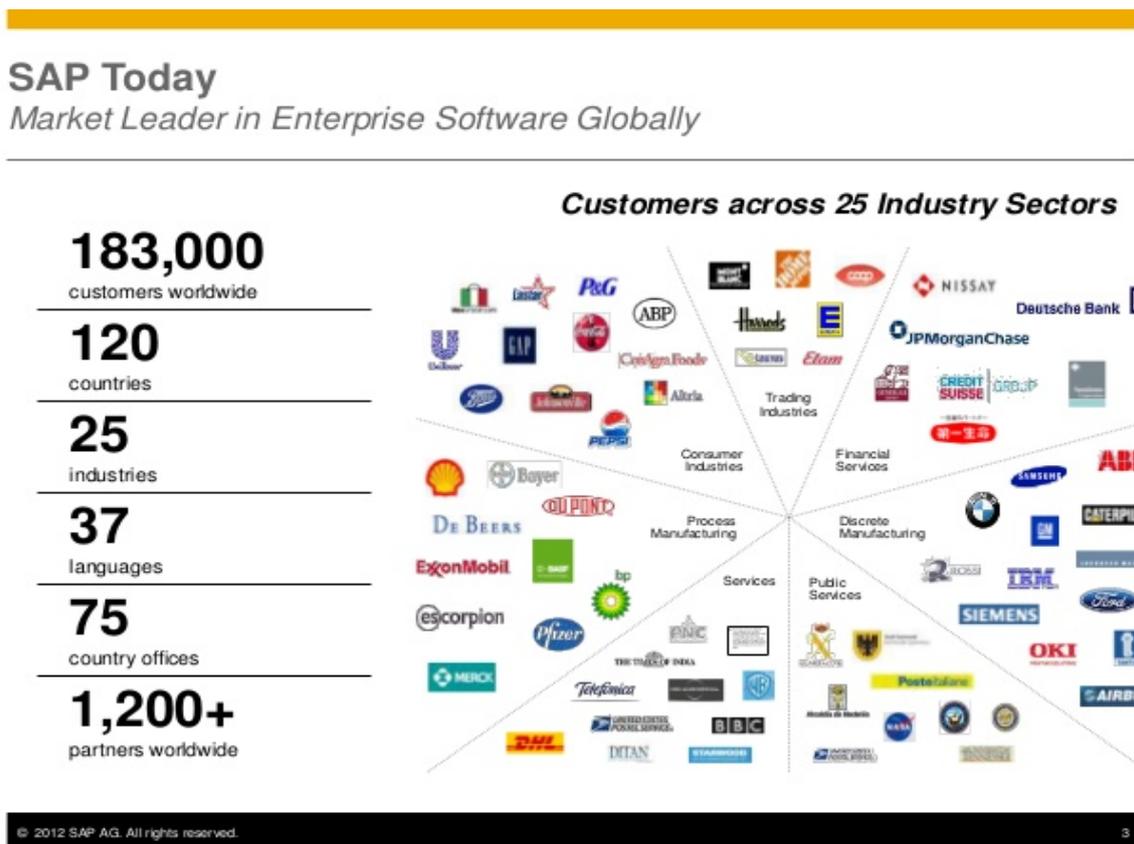


figure 2: SAP aujourd'hui(slideshare.com)

### 1.1.2 Industries

SAP offre des solutions qui couvrent les industries suivantes :

- L'énergie et les ressources naturelles :

Ce secteur englobe toutes ces industries :

Chimique, de transformation, d'exploitation minière, hydrocarbures, eau et énergie.

- Les services financiers. :

Ils comprennent les banques et les assurances.

- Les biens de consommation :

Ce sont les industries agroalimentaires, des biens de consommation, de mode, de distribution de détail et de commerce de gros.

- Les industries de production en discontinu :

Ce sont toutes les industries qui sont en relation avec l'aéronautique, l'automobile, les hautes technologies et les équipements industriels.

- Le secteur tertiaire :

Il comprend : le transport aérien, les travaux publics et bâtiment, les médias, les services professionnels, les sports et loisirs, les télécommunications et les voyages et transport.

- Les services publics :

Ils incluent la défense et la sécurité, les villes du futur, la santé, l'enseignement supérieur et la recherche et le secteur public.

### 1.1.3 Le portefeuille de solutions

Le portefeuille de solutions proposées par SAP comporte : (voir *figure 2*)

End-User Service Delivery					
Analytics	Strategic Enterprise Management	Financial Analytics	Operations Analytics	Workforce Analytics	
Financials	Financial Supply Chain Management	Financial Accounting	Management Accounting	Corporate Governance	
Human Capital Management	Talent Management		Workforce Process Management		Workforce Deployment
Procurement and Logistics Execution	Procurement	Supplier Collaboration	Inventory and Warehouse Management	Inbound and Outbound Logistics	Transportation Management
Product Development and Manufacturing	Production Planning	Manufacturing Execution	Enterprise Asset Management	Product Development	Life-Cycle Data Management
Sales and Service	Sales Order Management	Aftermarket Sales and Service	Professional-Service Delivery	Foreign-Trade Management	Incentive and Commission Management
Corporate Services	Real Estate Management	Project Portfolio Management	Travel Management	Environment, Health and Safety	Quality Management
SAP NetWeaver	People Integration	Information Integration	Process Integration	Application Platform	

*figure 3: portefeuille de solutions SAP (researchgate.net)*

- Les solutions ERP
- Les plateformes et solutions cloud
- Les solutions pour les achats
- Les outils d'analyse
- Les solutions CRM
- Les solutions IoT et Supply Chain
- Les ressources humaines
- Les solutions pour les finances

SAP offre trois types de déploiement de ses solutions, à savoir :

### Le déploiement de type auto-hébergement

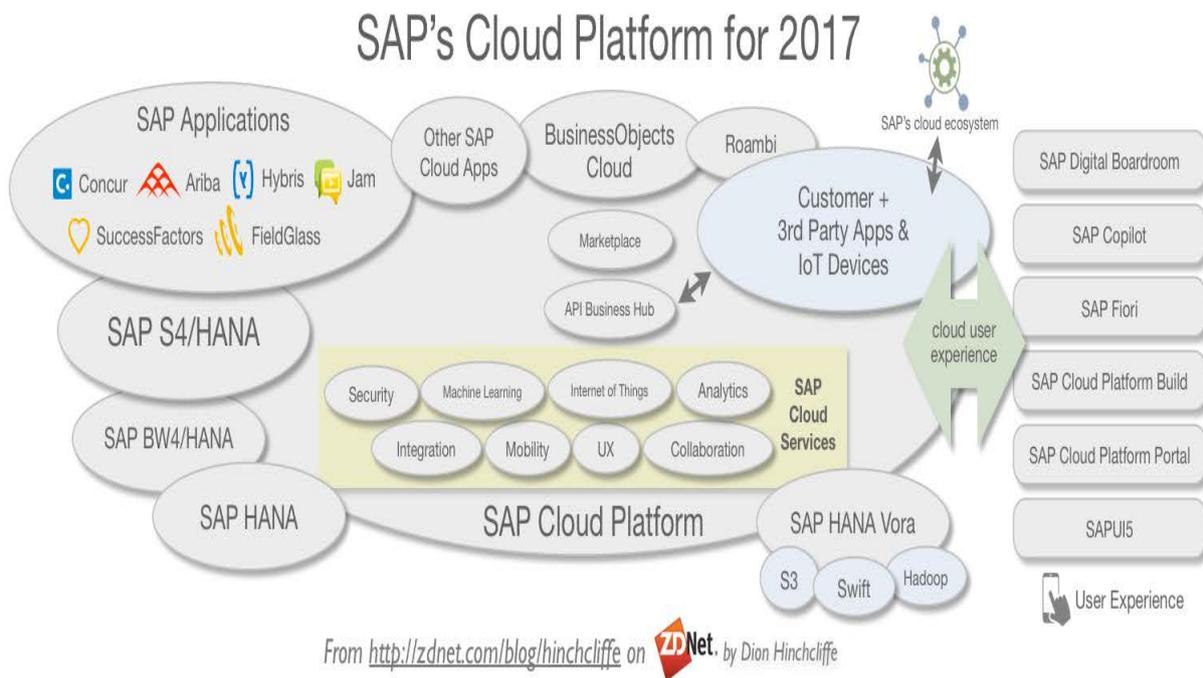
Les solutions sont installées sur les machines du client.

### Le déploiement de type Cloud

Les solutions sont hébergées sur le cloud. (figure 3)

### Le déploiement hybride

L'hébergement se fait avec les deux méthodes précédentes à la fois.



From <http://zdnet.com/blog/hinchcliffe> on ZDNet. by Dion Hinchcliffe

figure 4: la plateforme cloud de SAP(sap.com)

## **1.2 SONATRACH**

### **1.2.1 Présentation du groupe**

SONATRACH (Société nationale pour la recherche, la production, le transport, la transformation et la commercialisation des hydrocarbures) est une société algérienne, qui dispose d'un portefeuille d'activités assez riche, qui englobe toute la chaîne de valeur de l'industrie des hydrocarbures, ainsi que d'autres activités qui sont assurées par une partie de ses filiales.

Elle a été créée en 1963, sa première activité était le transport des hydrocarbures, mais grâce à une intégration progressive, elle touche actuellement toutes les activités des hydrocarbures et hors hydrocarbures, et elle compte aujourd'hui 154 filiales et participations, à l'échelle nationale et internationale.

La chaîne de valeur de SONATRACH se présente comme suit (de la partie amont à la partie aval) :

- L'exploration et la production :

Cette partie couvre l'exploration et l'exploitation des puits d'hydrocarbures, ainsi que toute activité de mise en place d'infrastructures nécessaires pour la partie amont.

- Le transport par canalisation :

Il comprend le transport du pétrole brut et du gaz brut aux raffineries et aux zones de transformation.

- La liquéfaction, le raffinage et la pétrochimie :

C'est au cours de cette phase, que le pétrole et le gaz bruts subissent des transformations, pour en extraire des dérivés, qui servent de produits à commercialiser.

- La commercialisation :

Elle forme la partie aval, en d'autres termes, la mise sur le marché des produits finis, destinés aux particuliers ou bien aux professionnels, à l'échelle nationale et internationale.

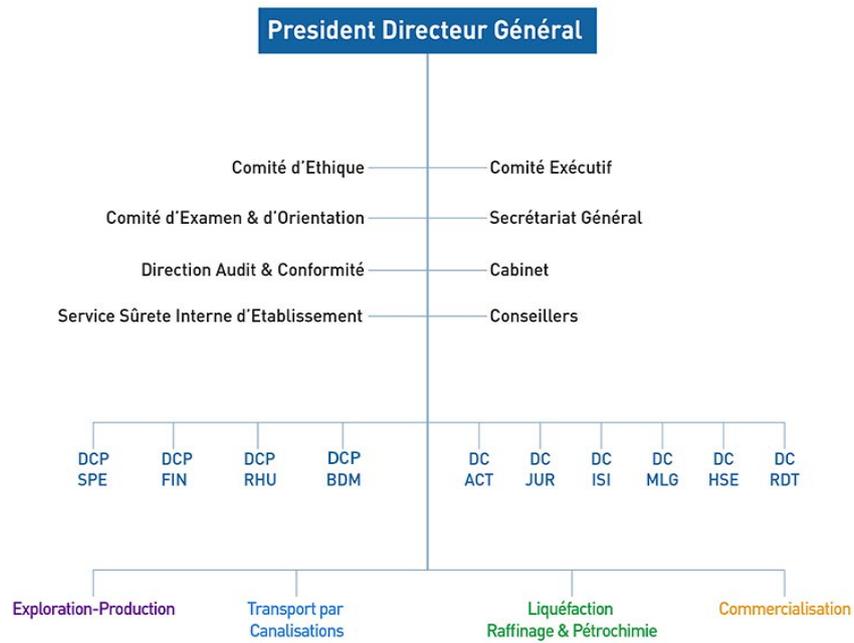


figure 5: Organigramme du groupe SONATRACH(sonatrach.dz)

Les filiales de SONATRACH sont organisées par holding, à savoir : (voir figure 5)

- Le holding SIP (société d'investissements et de participations)
- Le holding SPP (services para pétroliers)
- Le holding SVH (valorisation des hydrocarbures)
- Le holding SOALKIM (SONATRACH ALKIMAA)
- Le holding SGS (gestion et services de soutien)
- Le holding AIE (activités industrielles externes)

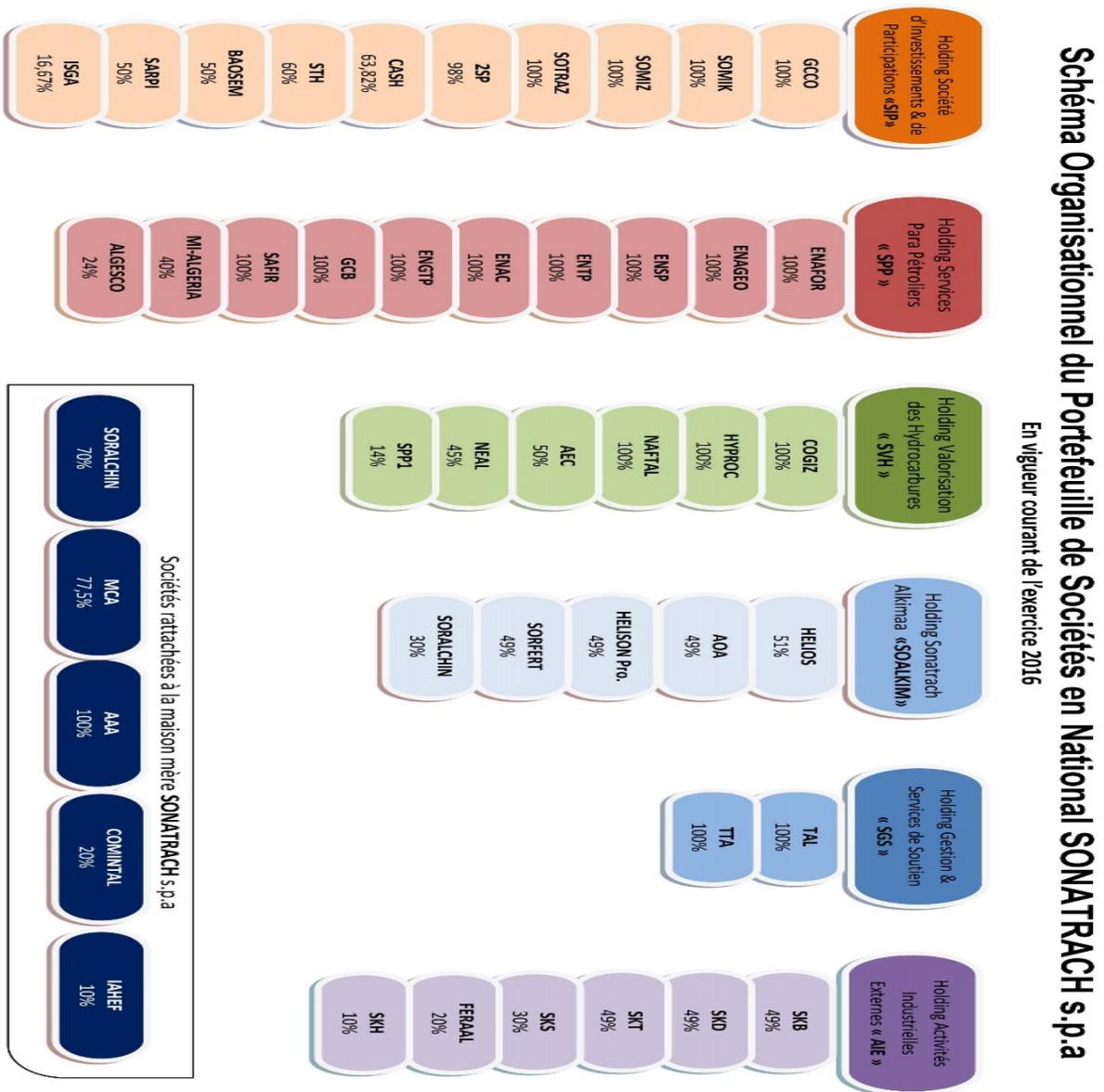


figure 6: les filiales du groupe SONATRACH par holdings(sonatrach.dz)

### 1.2.2 Filiales concernées par l'étude

Les filiales suivantes ont été choisies suite aux résultats de l'étude faite (voir chapitre 3), elles remplissent des critères qui leur donne la faveur d'être les premières à bénéficier de la transformation digitale.

#### 1.2.2.1 NAFTAL

NAFTAL offre des services et différents produits (dérivés de pétrole et de gaz naturel) pour deux catégories de clients, particuliers et professionnels.

##### Particuliers

- **Carburants terre**
  - Essence normale
  - Essence super
  - Essence super sans plomb
  - Gasoil
  - GPL/C
- **GPL**
  - Le butane commercial
  - Le propane commercial
- **Lubrifiants**
  - Les huiles pour moteurs à essence
  - Les huiles pour moteurs diesel
  - Les huiles pour transmission automobile
  - Les huiles spécialité automobile
- **Pneumatique** : commercialisation de pneumatiques de grandes marques qui touche les catégories suivantes :
  - Tourisme
  - Camionnette
  - Poids lourds
  - Industriel
  - Manutention
  - Agraire
  - Génie civil
- **Sirghaz**
- **Stations-service**

### Professionnels

- **Aviation**
  - Lubrifiants : huiles moteurs et turbines, fluides hydrauliques, graisse, compound, produits spéciaux et autres.
  - Carburant : JET A1, Kérosène et Avgas 100LL.
  
- **Marine**
  - Lubrifiants marine: huiles pour moteurs, cylindres, hydrauliques, engrenages et turbines, graisses et autre produits spéciaux.
  - Carburants marins : Gasoil et autres catégories de Fuel.
  
- **Bitumes**
  - Les bitumes modifiés aux polymères.
  - Les bitumes purs.
  - Les bitumes oxydés.
  - Les bitumes fluidifiés ou Cut-Backs.
  - Les émulsions de bitumes.
  
- **Produits spéciaux**
  - Les paraffines
  - Les cires
  - Les essences spéciales
  - Les solvants

#### 1.2.2.2 ENAFOR

ENAFOR est une filiale du groupe SONATRACH, elle est spécialisée dans les activités du forage, qui sont principalement le forage et le workover.

#### Services et activités

- Forage et workover :

Le forage est l'activité, par laquelle des gisements sont développés et exploités, afin d'extraire les fluides. L'entreprise effectue aussi le workover qui consiste en des opérations de reprise d'exploitation et d'entretien des puits.

- Le transport et déménagement des chantiers :

Cette activité est formée d'un ensemble d'opérations qui sont :

- Le DTM (démontage, transport et montage) des appareils et camps de forage
- L'approvisionnement des appareils en consommables
- L'approvisionnement en carburant et en pièces de rechange
- Le transport du matériel des chantiers
- Le transport du personnel des chantiers

- Maintenance pétrolière :

Cette maintenance touche les équipements et les installations de forage, la flotte de transport, cela par des remises en marche ou bien du Revamping.

Une GMAO a été mise en place, elle assure notamment la maintenance logistique, et la mise à jour d'une banque de données dédiée aux moyens.

- Hôtellerie et moyens communs :

C'est l'activité de prise en charge de l'hébergement, ainsi que de la restauration du personnel.

- Approvisionnements :

C'est le service qui assure l'alimentation des magasins et pièces de rechange, et en fournitures nécessaires à l'activité de l'entreprise.

### 1.2.2.3 HYPROC

HYPROC est une filiale de SONATRACH, spécialisée dans le transport maritime des hydrocarbures, des produits chimiques et du bitume, elle est dotée d'un savoir-faire dans le shipmanagement.

## 1.3 Le contexte de l'étude et de la problématique

Actuellement, SAP est en train de se développer en Afrique, pour acquérir de nouveaux marchés. Cependant, elle essaye de fournir des solutions efficaces, pertinentes et à forte valeur ajoutée, afin de pouvoir gagner une position stratégique dans ledit continent.

En 2017, SONATRACH a entamé une nouvelle stratégie qui a pour horizon 2030, cela en proposant une restructuration intégrale du groupe. Cette stratégie vise la modernisation de la compagnie et son alignement avec les bonnes pratiques internationales. L'une des problématiques qui ont été soulevées est celle de la qualité d'information. Il a été constaté que les workflows des filiales ne sont pas synchronisés, et cela a engendré une lourdeur de gestion et a empêché le bon déroulement des activités. Il est à souligner aussi, que les solutions dont disposent les filiales sont obsolètes, et ne répondent pas au réel besoin de leurs différents départements.

Pour remédier audit problème, le groupe a envisagé d'unifier son flux informationnel dans une seule base de données, en d'autres termes, utiliser un ERP assez puissant pour satisfaire ces besoins. Le groupe a opté pour les produits SAP, cela pour la raison que l'éditeur offre les meilleures solutions Oil & Gas sur le marché.

En Décembre 2017, Le groupe SONATRACH a fait l'acquisition d'un portefeuille de solutions SAP dédié à l'industrie des hydrocarbures, et il envisage d'aligner toutes ses filiales sur les mêmes pratiques de gestion et d'exploitation des systèmes d'information. Pour pouvoir bénéficier de la forte valeur ajoutée des solutions récemment acquises, la mise en place de ces dernières, doit être optimisée, de telle sorte à avoir le maximum de performance et d'évolution en un temps record. Cet objectif sera formalisé sous forme d'une feuille de route, qui sera établie en collaboration avec SAP et d'autres bureaux de conseils partenaires spécialisés en systèmes d'information.

### 1.4 Conclusion

Cet état des lieux était à la fois, une présentation des deux entreprises, ainsi que de leurs métiers. Il a permis de définir le contexte de l'étude et le réel problème à résoudre, et avec ces données de départ, on conclut ce chapitre. Le chapitre qui suit, va présenter l'état de l'existant, et va fournir les différentes connaissances, qui vont former le socle de la solution proposée.

# Chapitre 2

## Etat de l'art

---

### Introduction

L'état des lieux a montré les différents challenges et objectifs des deux entreprises, et a soulevé le cœur de la problématique à traiter. Ce chapitre a pour but de fournir par une revue de littérature, la terminologie, les concepts fondamentaux, ainsi que tous les rudiments des domaines des systèmes d'information, de l'intelligence artificielle et de la prise de décision. Ces définitions vont permettre par la suite, de bien expliquer les outils exploités pour la résolution de la problématique.

## 2.1 Système d'information

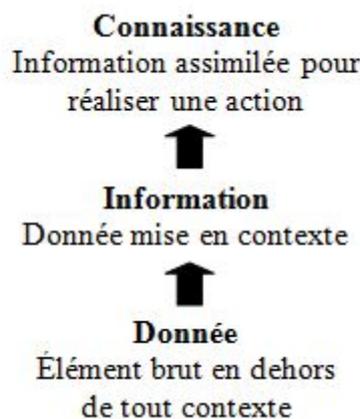
### 2.1.1 définitions

Le terme « système d'information » est composé de :

- **Système** : l'AFIS (association française d'ingénierie système) le définit comme un ensemble d'éléments en interaction en interne, et en externe avec un environnement. Il présente un certain nombre de caractéristiques qui le décrivent et qui le différencient des autres systèmes. Il peut être aussi décomposable en d'autres entités qui s'appellent « sous-systèmes ». (**Web c**)

- **Information** : cela provient du latin « informare » qui veut dire : donner des connaissances, des enseignements. Certains auteurs définissent l'information com suit : « l'information est le renseignement qui améliore la connaissance sur un sujet donné ». ( WILD & ROMAGNI, 1998)

L'information est le pont entre la donnée, qui est un élément brut dépourvu de tout contexte, et la connaissance qui représente une information exploitée pour déclencher une action. (voir *figure 7*)



*figure 8:le passage d'une donnée à une information(Web b)*

La notion « système d'information » a été revue plusieurs fois, et a fait l'objet de plusieurs définitions, qui diffèrent selon les contextes et les écoles, mais l'une des définitions les plus adoptée aujourd'hui, est celle d'O'Brien. (O'BRIEN, 2003)

O'Brien (O'BRIEN, 2003) définit un système d'information comme suit :

« Un ensemble de personnes, de procédures et de ressources qui recueillent de l'information, la transforment et la distribuent au sein d'une organisation. Et, aujourd'hui, avec le développement remarquable des outils informatiques, les SI utilisent l'informatique comme outil par excellence, pour accélérer tous les flux qui y circulent ».

### 2.1.2 types des SI

La typologie des systèmes d'information est fondée sur les couches décisionnelles au sein de la hiérarchie d'une entreprise.



figure 9: les couches décisionnelles au sein d'une entreprise(cgl-consulting.com)

Il existe trois types de systèmes d'information : (voir figure 7)

- **Les systèmes opérationnels (SO) :** appelés aussi les systèmes d'aide aux opérations; ces systèmes permettent de gérer les tâches récurrentes, et ils sont destinés principalement aux équipes chargées d'activités opérationnelles.

- **Les systèmes de management opérationnel (SMO) :**

Ces systèmes sont conçus pour répondre aux besoins des cadres dirigeants, en matière de gestion et de coordination des différentes tâches opérationnelles.

On en distingue deux catégories :

- Les systèmes de production de rapports : ils forment un support pour des décisions courantes, en se basant sur les données récoltées des opérations internes.

- Les systèmes d'aide à la décision : l'apport de ces systèmes, est leur capacité d'assister les décideurs, dans le cas des décisions semi-structurées ou mal structurées, cela en utilisant des technologies et des techniques telles que la fouille des données (data mining).

- **Les systèmes d'information stratégiques (SIS) :**

Cette catégorie de systèmes, permet de formuler et de simuler les décisions stratégiques, elle permet aussi de dérouler des scénarios pour fournir une vision plus claire, et de refléter l'impact à long terme de ce type de décisions assez importantes et critiques.

### 2.1.2 types de solutions informatiques pour la gestion des SI

Il existe de nombreuses solutions informatiques, qui assistent les systèmes d'informations des entreprises, elles visent principalement l'augmentation des performances sur le plan opérationnel et stratégique. Ces solutions technologiques sont :

- **ERP**

Acronyme de l'anglais du « enterprise resource planning » ou PGI pour « progiciel de gestion intégré », ce système offre une gamme de fonctionnalités qui regroupent l'essentiel des fonctions des entreprises, telles que la gestion des ressources humaines, la gestion des stocks, les ventes, les approvisionnements, la distribution, la finance et la comptabilité. Le fonctionnement de l'ERP repose sur le principe d'assurer une seule vérité, cela en centralisant les bases de données des différents départements en une seule. Chaque transaction dans ce système, sera synchronisée sur les différents départements. (EPINASSE, 2001)

- **CRM**

Acronyme de l'anglais de « customer relationship management », appelé aussi GRC pour « gestion de la relation client » ; cette solution offre la possibilité de gérer et de maintenir une bonne relation avec le client, cela par le biais d'une batterie d'outils, qui sont : le marketing ou « pre-sales », la vente, la gestion des services de la clientèle, les post-sales ou après-ventes. (EPINASSE)

- **SRM**

Ce terme vient de l'anglais « supplier relationship management », ou « gestion de la relation fournisseur » en français ; il représente un outil qui permet d'optimiser et de maintenir une bonne relation avec les fournisseurs, cela par le biais d'un ensemble de fonctionnalités, telles que les opérations de la sélection des fournisseurs, la facturation et le paiement. (EPINASSE, 2001)

- **SCEM**

De l'anglais « supply chain event management », qui veut dire la gestion des événements de la supply chain, cette solution dispose de plusieurs fonctionnalités, qui assistent les transactions logistiques, on cite : un système de pilotage, un système d'alarme, un système d'aide à la décision et un système de gestion documentaire. (EPINASSE)

- **MES**

Le MES est un système de contrôle de gestion, et du suivi des travaux en cours dans l'atelier. Un M.E.S (Manufacturing Execution System) conserve la trace de toutes les informations de fabrication en temps réel, et permet de recevoir des données en flux direct, à partir des systèmes de contrôle/commande, de supervision machine et des opérateurs. (**Web a**)

Le MES dispose d'un certain nombre de fonctionnalités (assises.club-mes.com, 2011), qui sont :

- Analyse des performances
  - Gestion des ressources,
  - Ordonnancement,
  - Cheminement des produits et des lots,
  - Gestion des documents,
  - Collecte et acquisition de données,
  - Gestion de la Qualité,
  - Gestion du procédé,
  - Gestion de la maintenance,
  - Traçabilité produit et généalogie,

• **APS**

Ce terme vient d'anglais de « Advanced Planning and Scheduling system », qui veut dire système de planification avancée, cet outil a la capacité d'intégrer des contraintes de l'entreprise, et de faire des planifications assez fiables. (voir figure 8)

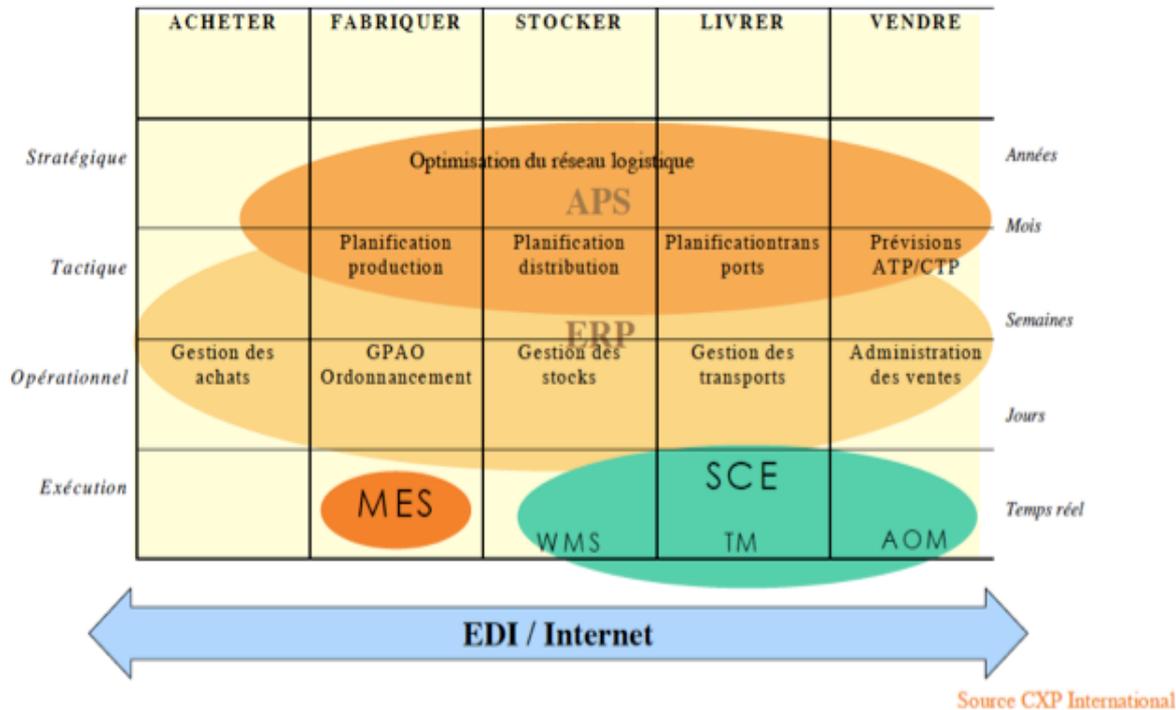


figure 10: les différentes solutions informatiques d'une entreprise en fonction des couches décisionnelles(supplychain-masters.fr)

• **BI (business intelligence)**

L'explosion des données en entreprises, représente aujourd'hui un énorme potentiel que chaque entreprise cherche à exploiter, pour en tirer un avantage concurrentiel ou pour en faire une base solide, pour rendre certains business models, autrefois bloqués par le manque de données, possibles et réalisables.

La prise de décision, utilise souvent comme support quelques indicateurs de l'activité de l'entreprise, appelés KPIs (key performance indicator), ces derniers sont réalisés suite à l'analyse, au tri et à l'exploration des données de l'entreprise.

La science qui offre la possibilité d'appliquer des analyses pertinentes et d'extraire la connaissance nécessaire du patrimoine d'informations dont dispose l'entreprise, est l'informatique décisionnelle. Le *Data Warehouse Institute* définit l'informatique décisionnelle comme suit :

« Les processus, les technologies et les outils nécessaires, pour transformer les données en informations, les informations en connaissance et la connaissance en plans, qui pilotent les actions profitables du business. L'informatique décisionnelle comprend l'entreposage de données, la *business analytics* et le management des connaissances ». (IBM Software Group, 2007)

La BI fonctionne selon une série d'opérations et de processus qui sont les suivants : (voir *figure 9*)

**ETL (extract, transform, load)** : après avoir repéré les différentes sources de données, les opérations de cette phase, vont assurer les tâches qui touchent la collecte, la consolidation et la structuration des données, qui seront par la suite introduites au datawarehouse.

**Le datawarehousing ou l'entreposage des données** : les données issues des opérations de l'ETL, seront hébergées dans des datawarehouses, selon une architecture dimensionnelle, et elles seront affectées vers les différents départements de l'entreprise sous forme de *datamarts* ou magasin de données.

**Le reporting** : cette phase, s'intéresse à la transformation des données du datawarehouse, en un support de prise de décision, cela par le biais des rapports et des tableaux de bord. Les technologies du data mining (la fouille des données) représentent également, un autre moyen pour tirer profit des données stockées dans le datawarehouse, et permettent d'avoir les connaissances cachées ou exprimées d'une manière non explicite.

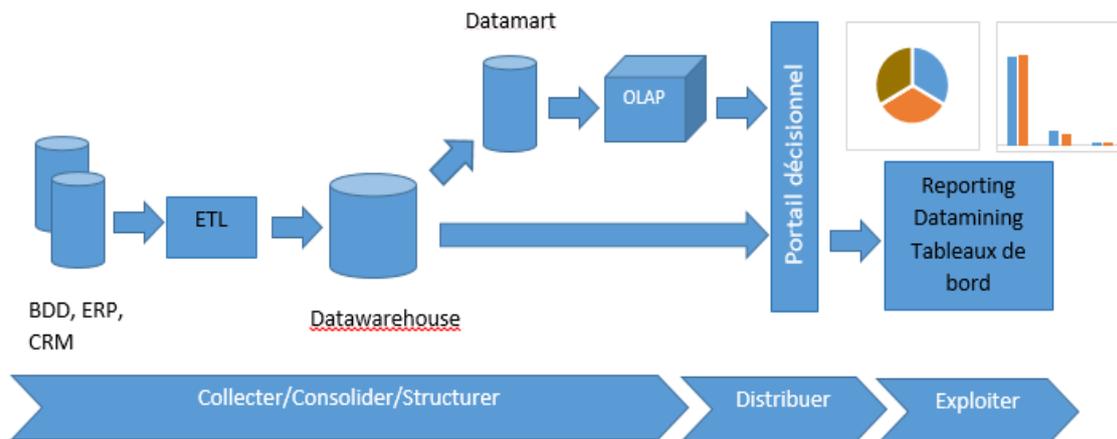


figure 11: le processus de la BI(pivovarov.org)

## 2.2 Intelligence artificielle

### 2.2.1 définitions

L'intelligence est la capacité d'un système à calculer, raisonner, percevoir des relations et des analogies, apprendre, stocker et récupérer les données d'une mémoire, résoudre les problèmes, comprendre des idées complexes, parler un langage couramment, classifier, généraliser et de s'adapter à des situations nouvelles.

Il existe plusieurs courants de pensées qui donnent des définitions différentes à l'intelligence artificielle. Ces définitions sont regroupées en quatre catégories, qui sont :

- L'intelligence artificielle vue comme « penser comme un humain » :

Cela veut dire les activités associées au raisonnement humain, des activités telles que la prise de décision et la résolution des problèmes.

- L'intelligence artificielle vue comme « agir comme un humain » :

Cette catégorie considère l'art de créer des machines qui réalisent des tâches qui nécessitent de l'intelligence, quand elles sont faites par des humains.

- L'intelligence artificielle vue comme « penser d'une manière rationnelle » :

Ce paradigme implique l'étude des facultés mentales, avec l'usage des modèles numériques.

- L'intelligence artificielle vue comme « agir d'une manière rationnelle » :  
L'étude du design des agents intelligents. (RUSSEL, 1994)

### 2.2.2 les sous-domaines de l'IA

L'intelligence artificielle a plusieurs champs de recherche ou sous-domaines, qui essaient de reproduire un ensemble de facultés intellectuelles de l'être humain, ces sous-domaines englobent :

- La modélisation des connaissances et le raisonnement automatique :

Cette sous-discipline inclut toute étude, qui vise le progrès des techniques de la modélisation et de la représentation des connaissances, étant donné que ces dernières sont souvent non disponibles d'une manière intégrale.

- La résolution des problèmes généraux :

Cette partie de recherche, s'intéresse principalement à la création de programmes ainsi que d'algorithmes capables de résoudre des problèmes réels.

- L'étude du langage naturel :

L'objectif de ce sous-domaine est de pouvoir effectuer une série de traitements, qui donnent la possibilité de la lecture ou de l'écriture du langage naturel.

- Machine Vision (vision artificielle) :

La vision artificielle est souvent reliée à la capacité d'une machine, de reconnaître des formes ou des profils à partir d'images ou de vidéos en entrée.

- Robotique :

Cette discipline, vise la production d'agents qui sont capables d'agir physiquement d'une manière intelligente, dans des situations données, tel que le ferait un être humain.

- L'apprentissage machine (machine learning) :

Ce sous-domaine représente l'ossature de l'outil proposé pour résoudre la problématique. Il sera détaillé dans la partie suivante.

## 2.3 Machine learning (apprentissage machine)

### 2.3.1 définitions

L'apprentissage machine est une discipline en informatique, qui concerne l'intelligence artificielle. Elle s'est développée avec les études de la reconnaissance des formes et avec l'évolution des modèles numériques d'intelligence artificielle. Cette science vise le développement d'algorithmes qui sont capables d'apprendre et de faire des prévisions, à partir des données qui leur sont injectées comme inputs. La qualité de ces algorithmes est de pouvoir générer des modèles à partir des données en entrée. Cela permet de prendre des décisions ou de faire des prévisions, en se basant sur les données et non pas sur des instructions rigides et strictes.

Le domaine de l'apprentissage machine, est fortement relié au domaine des statistiques, une discipline mathématique qui peut être orientée prévisions. Des exemples concrets de l'application de l'apprentissage artificiel sont: le détecteur de spam, le système OCR (optical caractere recognition) de reconnaissance optique des caractères, les moteurs de recherche tels que Google et Bing et la vision machine. L'apprentissage machine est souvent confondu avec le fouillage de données ou *data mining* de l'anglicisme. Le premier s'intéresse aux analyses prédictives, par contre le deuxième est orienté exploration de données et de connaissances. (RUSSEL, 1994)

### 2.3.2 les types d'apprentissages

Il existe plusieurs problématiques et tâches en apprentissage machine, qui couvrent un certain nombre d'aspects pratiques. Les spécialistes les classent en trois catégories principales :

- L'apprentissage supervisé

Cette classe est caractérisée par le fait que le type d'entrées et de sorties sont préalablement définies par un « maître », ensuite le but est de générer des modèles qui satisfont les entrées et les sorties qui ont été déterminées au départ.

- L'apprentissage par renforcement

Le paradigme qui couvre cette catégorie, est le besoin d'atteindre un objectif défini, en interagissant avec un système dynamique, et sans la présence explicite d'un maître qui pilote le système,

- L'apprentissage non-supervisé

Contrairement à l'apprentissage supervisé, l'apprentissage non-supervisé ne repose pas sur la définition initiale des entrées, mais il est fait de telle sorte à pouvoir bâtir des modèles indépendamment des inputs. Cette classe d'algorithmes traite en majeure partie les problèmes de regroupement ou de « *clustering* ».

Le clustering forme le cœur de la problématique traitée dans cette étude, de ce fait la qualité des résultats obtenus sera tributaire du bon choix d'algorithme.

## 2.4 Clustering

### 2.4.1 définitions

Le clustering est une procédure dans l'analyse multivariée. Il est établi pour explorer la structure inhérente des données, où les données dans le même cluster sont les plus semblables possible, et elles sont les plus différentes possibles avec les autres clusters.

L'être humain opère dans des situations données, en utilisant le principe du clustering. Il peut par exemple distinguer les océans des continents sur une carte géographique. La notion du clustering est souvent sujette à confusion avec la notion du classement, cette dernière consiste à affecter des objets à des groupes ou classes déjà existants. Par contre le clustering crée les groupes et ne les obtient pas des conditions initiales.

Le clustering est une analyse exploratrice des données. Il peut avoir un minimum d'informations sur les paramètres des clusters résultants ou bien aucune. Ces paramètres sont :

- Le nombre de clusters
- La position relative ou absolue des clusters
- La taille des clusters
- La forme des clusters
- La densité des clusters

### 2.4.2 la norme de Minkowski

Pour mesurer la similarité des objets des différents groupes de données, de différentes distances sont envisageables ; et ces dernières sont généralisées dans une seule norme qui s'appelle la norme de « Minkowski ».

La norme de Minkowski mesure la distance  $d$  entre deux objets  $x$  et  $y$ , cela en considérant la différence entre leurs différentes composantes, elle est définie comme suit :

$$d(x,y) = L_q(x,y) = \sqrt[q]{\sum_{i=1}^n (x_i - y_i)^q}$$

Avec :

$d(x,y)$  : distance entre  $x$  et  $y$

$L_q(x,y)$  : distance de Minkowski entre  $x$  et  $y$

$q$  : un exposant donné.

$i$  : l'indice de la composante de la variable.

Pour des valeurs spéciales de l'exposant «  $q$  » on distingue deux types de distances les plus communément utilisées :

$q=1$  :

La norme dans ce cas de figure est appelé **la distance de Manhattan** et elle est définie comme suit :

$$d(x,y) = L_1(x,y) = \sum_{i=1}^n |x_i - y_i|$$

$q=2$  :

La norme devient pour cette valeur de «  $q$  », **la distance euclidienne** et elle est définie comme suit :

$$d(x,y) = L_2(x,y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

### 2.4.3 les algorithmes du clustering

Il existe de nombreux algorithmes de clustering, avec différentes approches. Ils sont classés en deux familles :

#### 2.4.3.1 le clustering hiérarchique

Cette famille de méthodes impose une structure hiérarchique aux différents objets et à leurs clusters. La première extrémité comporte un seul cluster qui regroupe tous les éléments, et pour l'autre extrémité, chaque élément est un cluster. Par conséquent si l'on souhaite obtenir un nombre  $k$  de clusters, la hiérarchie doit être parcourue en profondeur.

Le clustering peut être effectué d'une manière clivante. Cela veut dire qu'à partir d'un seul cluster vers les  $k^{\text{ième}}$  clusters, et il peut être également élaboré d'une manière agglomérée cela en partant du plus grand nombre de cluster vers les  $k^{\text{ième}}$  clusters. La deuxième méthode est plus favorable car elle est plus intéressante en terme de complexité algorithmique

### L'algorithme du clustering hiérarchique aggloméré

#### Entrées :

- Un ensemble d'objets  $\mathbf{O} = \{\mathbf{o}_1, \dots, \mathbf{o}_n\} \subseteq \mathbb{R}^m$
- Une fonction de distance définie comme suit  $d : \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}$

#### Traitement :

1. Pour tout objet  $\mathbf{o}_i \in \mathbf{O}$  faire :
  - i. Créer le cluster  $\mathbf{C}_i = \{\mathbf{o}_i\}$
2. On définit  $\mathbf{C} = \{\mathbf{C}_1, \dots, \mathbf{C}_n\}$
3. Tant que  $|\mathbf{C}| \neq 1$  faire :
4. Pour toute paire de clusters  $\langle \mathbf{C}_i, \mathbf{C}_{j \neq i} \rangle \in \mathbf{C} \times \mathbf{C}$  faire
  - i. Calculer  $d(\mathbf{C}_i, \mathbf{C}_j)$
  5. On définit la fonction  $\text{choix}(\mathbf{C}_i, \mathbf{C}_j) = \forall \langle \mathbf{C}_{k \neq i}, \mathbf{C}_{l \neq k, j} \rangle \in \mathbf{C} \times \mathbf{C} : [d(\mathbf{C}_i, \mathbf{C}_j) \leq d(\mathbf{C}_k, \mathbf{C}_l)]$
  6. Pour  $\text{choix}(\mathbf{C}_i, \mathbf{C}_j)$  faire :
    - On définit  $\mathbf{C}_{ij} = \mathbf{C}_i \cup \mathbf{C}_j$
    - On définit  $\mathbf{C}' = \mathbf{C} \setminus \{\mathbf{C}_i, \mathbf{C}_j\}$
    - $\mathbf{C} = \mathbf{C}' \cup \mathbf{C}_{ij}$

**Fin.**

Cet algorithme inclut la notion de distance entre clusters, ce paradigme est traité par plusieurs types de distances qui sont :

- **Le plus proche cluster voisin**

La distance entre deux clusters, est définie comme étant la distance minimale entre les objets des clusters, comme le montre l'équation suivante :

$$d(\mathbf{C}_i, \mathbf{C}_j) = d_{\min}(\mathbf{C}_i, \mathbf{C}_j) = \min_{x \in \mathbf{C}_i, y \in \mathbf{C}_j} d(x, y)$$

- **Le plus loin cluster voisin**

Dans ce deuxième cas de figure, la distance entre deux clusters, est la plus grande distance qui sépare les objets des clusters, tel que le montre cette formule :

$$d(\mathbf{C}_i, \mathbf{C}_j) = d_{\max}(\mathbf{C}_i, \mathbf{C}_j) = \max_{x \in \mathbf{C}_i, y \in \mathbf{C}_j} d(x, y)$$

- **La distance entre les centroïdes des clusters**

Le centroïde d'un cluster est défini comme étant le barycentre de ses objets. La distance dans

ce cas, est définie comme étant la distance entre les centroïdes de deux clusters, cette formule illustre ce principe :

$$d(C_i, C_j) = d_{bary}(C_i, C_j) = d(cen(C_i), cen(C_j))$$

Avec  $cen(C)$  le centroïde d'un cluster  $C$ .

- **La distance moyenne entre clusters**

Cette distance est la moyenne des distances entre les différents objets des deux clusters. L'équation de cette distance est la suivante :

$$d(C_i, C_j) = d_{moy}(C_i, C_j) = \frac{1}{|C_i| * |C_j|} * \sum_{x \in C_i} \sum_{y \in C_j} d(x, y)$$

### 2.4.3.2 le clustering de partitionnement

Ce deuxième type de clustering, repose sur le principe de partitionner les objets de l'ensemble des données, en des clusters distincts, et on parle de « hard clustering », ou bien en clusters qui peuvent avoir des éventuels éléments en commun, le clustering est décrit dans ce cas de « soft ». Les algorithmes de clustering démarrent avec une partition initiale des clusters, qui sera améliorée au fur et à mesure, jusqu'à l'obtention d'une partition finale stationnaire. Les clusters initiaux peuvent être générés aléatoirement.

#### L'algorithme des K-moyennes

L'algorithme du clustering par les k-moyennes est une méthode non-supervisée du hard clustering, qui affecte  $n$  objets  $o_1, \dots, o_n$  vers un nombre  $k$  prédéfini de clusters  $C_1, \dots, C_k$ . Les objets sont initialement affectés vers des clusters, en d'autres termes vers les centroïdes les plus proches ; ensuite ces derniers seront réajustés et recalculés. La condition pour achever le déroulement de l'algorithme est la stationnarité des clusters. (voir figure 120)

#### L'algorithme

##### Entrées:

- Un ensemble d'objets  $O = \{o_1, \dots, o_n\} \subseteq \mathbb{R}^m$
- Une fonction de distance définie comme suit  $d : \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}$
- Un partition de  $k$  clusters (qui peuvent être aléatoire) :  $C = \{C_1, \dots, C_k\}$

##### Traitement:

1. Pour tout cluster  $C_i \in C$  faire :  
Calculer le centroïde du cluster  $cen_i \subseteq \mathbb{R}^m$
2. Pour tout objet  $o \in O$  faire :  
Pour tout cluster  $C_i \in C$  faire :  
Calculer  $d(o, C_i) = d(o, cen_i)$   
On définit  $choixCls(o, C_0) : \forall C_j \in C : [d(o, cen_{C_0}) \leq d(o, cen_{C_j})]$
3. On définit  $C'$
4. Pour tout objet  $o \in O$  faire :

Pour  $choixCls(o, C_0)$  faire :

Affecter  $o$  à  $C'_0$

5. Si  $C \neq C'$  alors  $changer = vrai$

Sinon  $changer = faux$

6. Fin si  $changer = faux$

Fin.

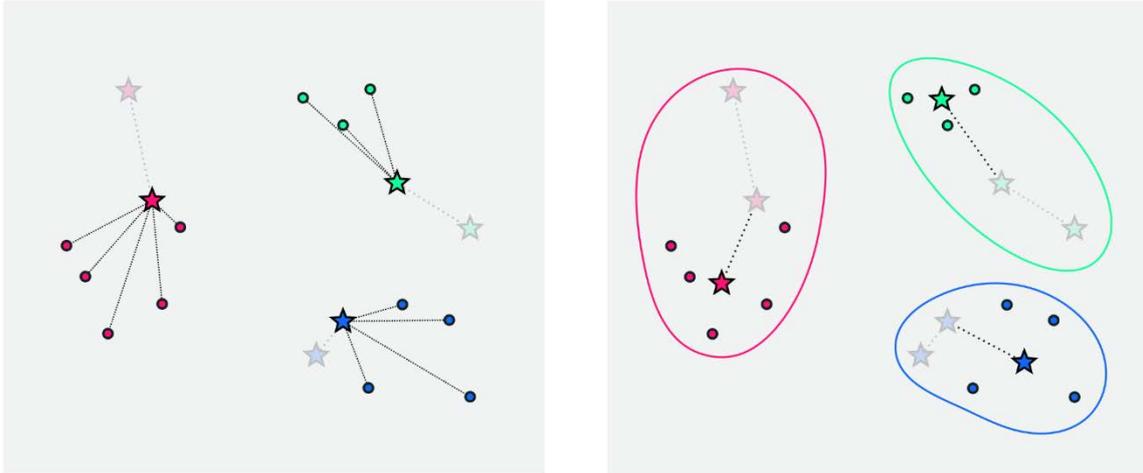


Figure 10: l'évolution des clusters et de leurs centroïdes dans l'algorithme des k-moyennes(github.io)

## 2.5 Aide multicritère à la décision

### 2.5.1 définitions

L'aide à la décision est une discipline, à laquelle ont recours les industriels et les managers quotidiennement, afin de résoudre des problématiques de différentes natures (planification, approvisionnement, stratégie ...), cela en étant soumis à des contraintes de types variées aussi (temps, coûts, qualité ...).

L'auteur B. Roy (ROY, 1985) définit l'aide à la décision comme suit :

« L'activité de celui qui, prenant appui sur des modèles clairement explicités mais non nécessairement complètement formalisés, aide à obtenir des éléments de réponse aux questions que se pose un intervenant dans un processus de décision ».

L'aide à la décision est souvent décrite de « multicritère », compte tenu du fait qu'elle soit souvent soumise à des contraintes multiples, cela étant une conséquence de l'environnement complexe auquel font face les décideurs.

### 2.5.2 les types de problématiques

- Le choix :

Cette problématique consiste à faire la sélection d'un sous-ensemble d'actions parmi un ensemble d'actions potentielles.

- Le tri :

Cette catégorie est similaire à un problème d'affectation, où des actions seront attribuées à des catégories données comme référence.

- Le rangement ou le classement :

Le problème traité dans ce cas de figure est la nécessité de créer des classes d'actions, ainsi que de les ordonner.

- La description :

Ce type s'intéresse à la description générale des conséquences des décisions, afin de permettre au décideur de mesurer l'impact et d'être plus attentif sur la démarche à entreprendre.

### 2.5.3 l'aide à la décision et l'apprentissage

Les techniques d'apprentissage (supervisé et non-supervisé) comptent parmi les outils utilisés pour résoudre les problèmes d'aide multicritère à la décision.

Les algorithmes du clustering de l'apprentissage non-supervisé ont la qualité de pouvoir regrouper des entités de composantes multiples dans des classes ou clusters bien illustratifs. Par conséquent, ils sont utilisés pour résoudre des problématiques de type « classification », dans le cadre de l'aide multicritère à la décision.

## 2.6 Conclusion

Ce chapitre était une synthèse des bases ainsi que des notions indispensables à la modélisation et à la résolution de la problématique donnée. Il va jouer un rôle significatif quant à la compréhension de ce qui sera présenté dans le prochain chapitre.

La problématique à résoudre exige d'avoir comme résultat, des regroupements de solutions SAP bien définis, et non inter-reliés, en partant de référentiels préalablement établis. Etant donné cet ensemble de contraintes, l'algorithme des k-moyennes semble être le plus approprié pour répondre à tous ces besoins.

Les parties du chapitre suivant seront majoritairement axées sur les points suivants :

- La présentation des différentes données
- L'application de l'algorithme d'apprentissage machine
- L'exploitation de l'outil proposé
- La présentation des résultats obtenus.

# Chapitre 3

## Développement de la solution

---

### **Introduction :**

Dans le chapitre précédent, un ensemble d'outils quantitatifs et qualitatifs ont été présentés afin de créer une base théorique pour ce travail. Ce chapitre va mettre l'accent sur l'application des techniques d'un ensemble de disciplines, dans le but de résoudre une problématique réelle, d'ordre stratégique pour le client de SAP.

Le groupe SONATRACH s'attend à obtenir une feuille de route qui oriente la mise en place des solutions SAP, autrement dit, un plan d'action qui donne une classification des différentes solutions par filiale. Cela, en suivant une logique de priorisation axée sur la création de valeur, de la manière la plus optimale possible.

La feuille de route est le résultat d'une classification basée sur le clustering, dans un premier temps sur les filiales pour déterminer les plus prioritaires, ensuite sur l'état de leurs différents systèmes d'information, afin de définir les solutions informatiques les plus pertinentes, mais aussi les plus urgentes et importantes.

Le processus de la mise en place de la feuille de route a été une succession d'étapes, qui sont :

- Le développement du logiciel qui construit le clustering
- La collecte des données
- L'analyse des résultats
- La mise en place du plan d'action

### 3.1 Les étapes du développement de la solution

#### 3.1.1 Développement du logiciel de clustering

La nécessité du développement d'une solution automatisée est liée à la complexité de la problématique. Il existe d'une part un périmètre d'intervention très large (nombreuses filiales, périmètre fonctionnel très important) et d'autre part un portefeuille de solutions pouvant prendre en charge les besoins des entreprises du groupe SONATRACH également très important.

Le développement du logiciel a nécessité les étapes suivantes :

##### 3.1.1.1 Définition des besoins

Cette étape est primordiale, car elle répond à un ensemble de questions qui concernent le projet, notamment la faisabilité, cela en étudiant le concept et les résultats attendus, et les faire correspondre aux compétences et aux moyens disponibles, et suite à cette opération toutes les structures algorithmiques et les architectures adéquates peuvent être conçues et exploitées.

Les besoins de cette étude sont :

Un programme capable de mettre en œuvre l'algorithme des k-moyennes sur des données disponibles dans des fichiers plats ou MS EXCEL, et pouvoir par la suite afficher les résultats sur un plan avec une perte d'information la plus minimale possible. Il doit donner également la possibilité de sauvegarder les résultats sous forme de rapports, et afficher un graphique montrant la distribution finale des points et des clusters.

##### 3.1.1.2 Modélisation de l'architecture

Après avoir étudié les besoins et défini les outils algorithmiques adéquats, l'architecture suivante a été mise en place (figure 11) :

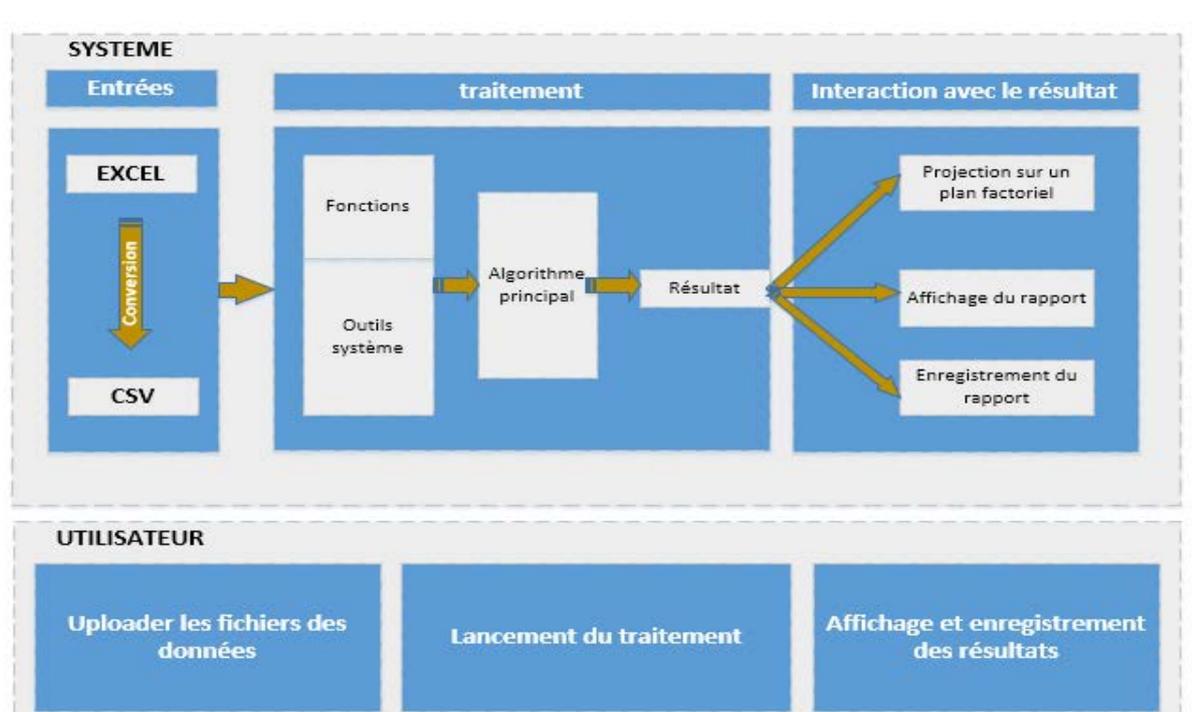


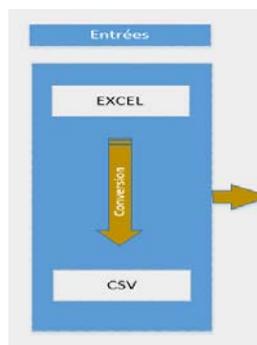
figure 13: l'architecture du programme informatique développé

La partie machine représente tous les traitements et les opérations qui sont invisibles à l'utilisateur, et la partie utilisateur englobe l'interface graphique et HMI (*human machine interface*, ou interface homme-machine), et grâce à elle l'utilisateur peut interagir et lancer les traitements désirés sur les données.

### La couche système

- Les entrées

Cette partie permet d'insérer les fichiers qui comportent les données à exploiter, ceux-là étant sous format EXCEL seront convertis en format CSV, car ce dernier simplifie la récupération de la matrice de données qui sera traitée. Le nombre de clusters représente aussi une entrée importante, car il permet de vérifier que le fichier des centroïdes initiaux correspond effectivement à ce nombre de clusters. L'utilisateur aura la possibilité de déclarer ces entrées à l'aide de boutons et de champs à remplir. (voir *figure 12*)



*figure 12: les entrées*

- Le traitement

Cette partie représente l'exploitation des données en entrée, cela en leur appliquant l'algorithme des k-moyennes. Ce dernier est en interaction permanente avec d'autres fonctions programmées, ou tirées directement des bibliothèques du système. Par conséquent, un résultat sera généré et communiqué sous différentes formes, qu'offre la couche « interaction avec le résultat ». Cette partie représente une boîte noire pour l'utilisateur et il n'en a pas les droits d'accès. (voir *figure 13*)

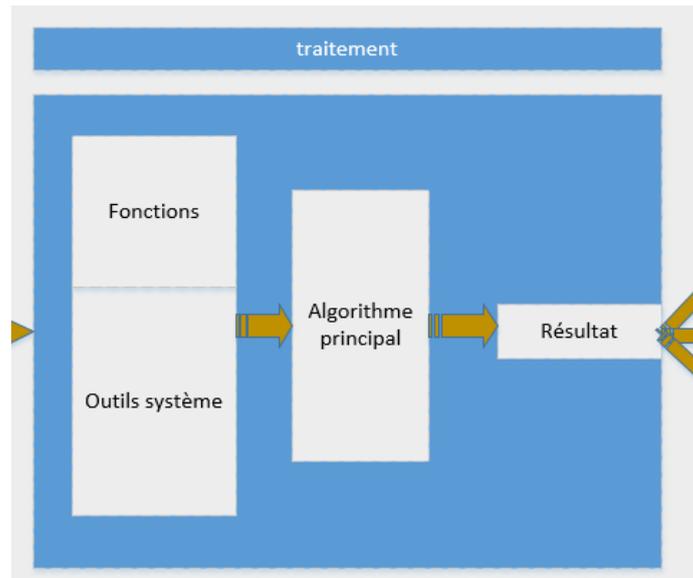


figure 14: le traitement

- L'interaction avec le résultat

Le résultat des étapes précédentes est un ensemble de clusters, ces derniers pouvant être visualisés par un rapport généré automatiquement suite à l'exécution de l'algorithme. La visualisation graphique est également possible. Elle montre la projection des données et des résultats sur un plan factoriel. Ce dernier est généré à partir de l'analyse factorielle des données initiales. (voir figure 14)

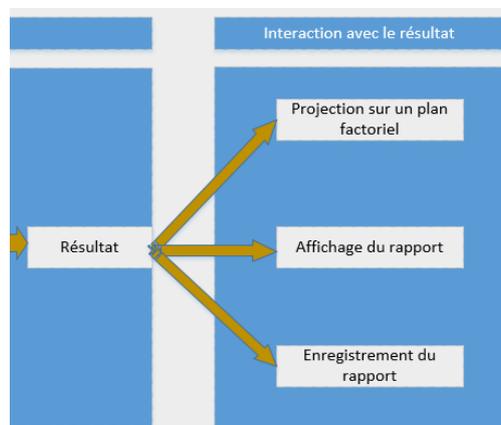


figure 14: les résultats ou sorties

### La couche utilisateur

Cette partie de l'architecture comporte tous les outils visuels (tels que les boutons et les champs à remplir), afin d'accélérer l'entrée des données et faciliter l'utilisation de l'application. L'utilisateur a la possibilité d'effectuer les opérations suivantes : l'insertion des données, le lancement du traitement et l'affichage ou l'enregistrement des résultats. (voir figure 15)



figure 15: la couche "utilisateur"

### 3.1.1.3 Codage et programmation

L'architecture a permis de définir un ordre et une approche avec laquelle doit être lancée la programmation. Le choix du langage Python est justifié par plusieurs facteurs qui sont :

- La disponibilité de la compétence
- La disponibilité de bibliothèques qui disposent d'outils qui répondent au cahier des charges.
- Une référence pour la programmation orientée *data science* (la science de l'analyse des données), cela par sa disposition d'outils algorithmiques favorables pour la gestion et l'analyse des données
- La simplicité, l'efficacité
- La possibilité de développement d'interfaces graphiques avec le langage, sans avoir recours à d'autres outils.

La programmation a été entamée sur la partie système, les outils d'acquisition d'entrées, sont programmés en premier, avant de passer aux fonctions et à l'algorithme des k-moyennes, et enfin le résultat. Après avoir achevé la partie système, cette dernière a été intégrée dans la partie utilisateur, pour permettre aux outils graphiques d'y avoir accès et de donner la main à l'utilisateur.

Le code source est disponible dans l'annexe 2.

### 3.1.1.4 la validation

La validation du logiciel a été faite par des jeux de test obtenus librement sur des plateformes dédiées, cela en ayant un ensemble de données et les résultats de leur clustering. A l'issue de ces tests, il était possible de vérifier la fiabilité du programme.

Cette partie met fin à l'étape de la programmation, et sort avec le résultat qui est un programme capable d'effectuer le clustering

. Ci-dessous l'interface finale :

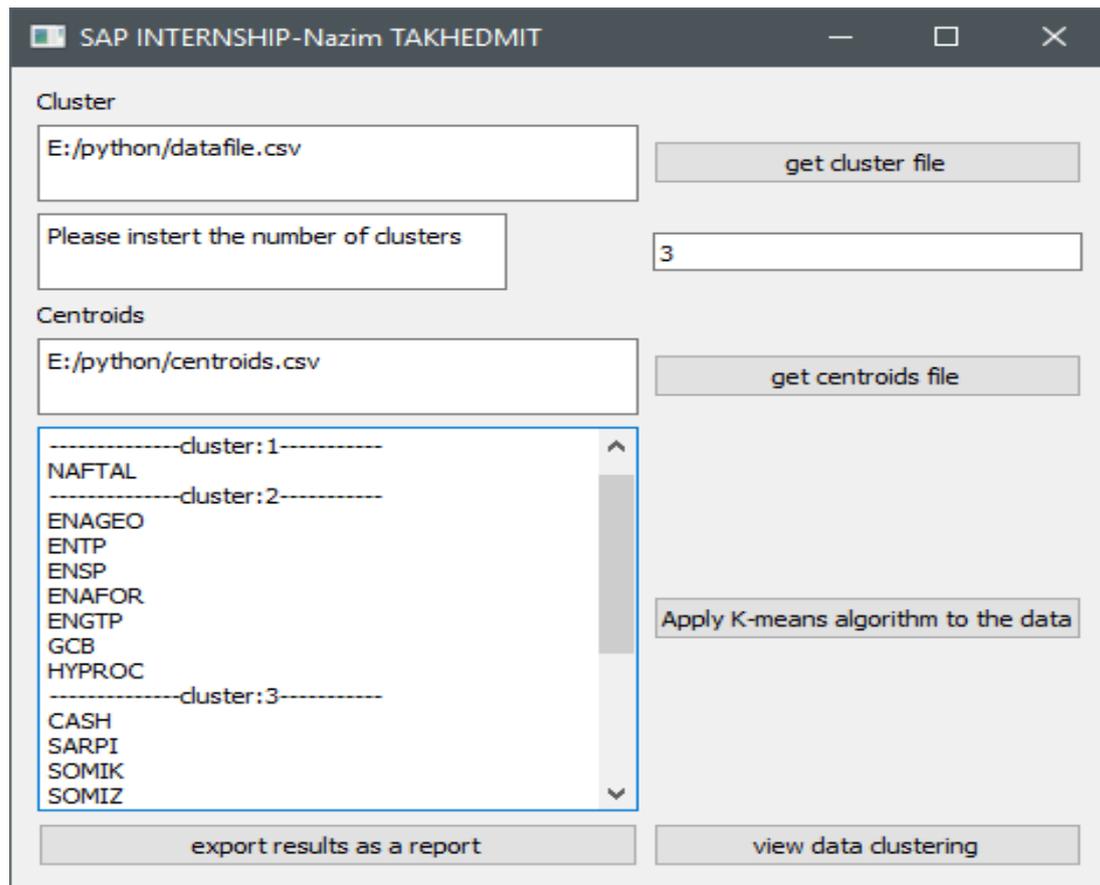


figure 16: interface graphique principale

### 3.1.2 la recherche des données

Les résultats de la feuille de route prennent en considération, les filiales du groupe ainsi que les solutions du portefeuille SAP acquis.

Les données sur les filiales ont été choisies, en se basant sur la création de valeur, ainsi que sur des contraintes qui influencent la mise en place des solutions SAP. Par rapport aux données relatives à la maturité des systèmes d'information des filiales, un questionnaire a été adressé aux directions des systèmes d'information de quelques filiales. Ce questionnaire est fondé sur un autre questionnaire de référence, qui est utilisé par SAP pour évaluer la maturité et les besoins en systèmes d'information des entreprises.

#### 3.1.2.1 les données sur les filiales

Suite à des réunions avec des experts SAP, les types de données à chercher sont définis, à savoir :

- Le chiffre d'affaire : car il reflète l'activité et l'importance économique et commerciale de la filiale
- Le résultat : il est choisi pour sa corrélation directe avec la valeur ajoutée créée pour le groupe
- Le nombre de salariés : cela traduit le niveau d'importance et de complexité de l'entreprise ainsi que le nombre d'utilisateurs potentiel des solutions à mettre en place

- Le nombre d'années d'activité : ce chiffre peut être exploité pour apprécier la maturité d'une entreprise dans son propre domaine d'activité.

Suite à des recherches de type data-crunching et aussi après avoir contacté certaines filiales pour complément et validation, il a été possible d'obtenir des données sur les différentes filiales du groupe, et qui sont résumées ci-dessous :

Filiale	Chiffre d'Affaire (MDA)	Résultat Net (MDA)	Effectif (agents)	Année de création
CASH	9.000	1.000	554	1999
SARPI	3.603	952	1.412	1992
SOMIK	1586	-1,2	1.936	1991
SOMIZ	5.000	350	2.400	1991
SOTRAZ	1.824	102	1.208	1991
GCCO	1.120	79	476	2011
2SP	5.105	11	6.215	1997
STH	3.761	926	750	2014
BAOSEM	282	91	743	2002
ENAGEO	18.777	1.266	6.307	1981
ENTP	53.429	12.309	7.567	1983
ENSP	18.366	3.480	2.904	1981
SAFIR	488	82	159	1991
ENAFOR	32.945	6673	7.567	1983
ENAC	12.294	2224	3.133	1981
ENGTP	24.557	1.264	10.486	1980
GCB	24.278	6.371	9.081	1981
NAFTAL	332.739	8.193	30.709	1981
COGIZ	789	319	73	1998
HYPROC	19.633	896	1.729	1982
TAL	8.179	823	1220	1998
HELIOS	2622	1399	51	1991

*Tableau 1: données des différentes filiales (données de l'année 2017)*

### 3.1.2.2 les données sur les SI des filiales

En s'inspirant d'une méthodologie de référence de SAP (*Franchise for Success*), qui contient des questions ciblées par secteur d'activité, un questionnaire a été réalisé et comporte des questions très spécifiques et ciblées, qui permettent d'avoir des données très pertinentes sur l'état du SI d'une entreprise. Le client aura à répondre au questionnaire, en attribuant un score (basé sur une échelle de 1 à 8), à chaque activité.

Les questions que le formulaire contient sont en Annexe 3.

Les périmètres fonctionnels ou départements visés dans chaque question sont :

- Finance
- Ressources humaines
- Recherche et développement
- Ingénierie
- Logistique
- La gestion des projets et du portefeuille
- La gestion des bases de données
- La gestion des actifs
- La gestion des arrêts et des pannes (maintenance)
- Hygiène et sécurité

Le choix de ces solutions est basé sur le type de solutions existant dans le portefeuille SAP de l'industrie OIL & GAS. Les filiales adressées sont celles qui sont identifiées comme prioritaires selon les résultats de la première étude de clustering. Comme indiqué précédemment, ces filiales sont : NAFTAL, HYPROC et ENAFOR.

Les résultats obtenus dans cette étape sont :

(Les chiffres dans les colonnes renvoient aux questions, selon l'ordre précédemment indiqué)

NAFTAL							
	1	2	3	4	5	6	7
<b>Finance</b>	4	3	6	4	7	5	2
<b>Ressources humaines</b>	4	4	5	3	5	5	2
<b>Recherche et développement</b>	1	1	1	1	1	2	1
<b>Ingénierie</b>	2	2	2	3	3	2	2
<b>Logistique</b>	3	3	2	3	5	4	3
<b>La gestion des projets et du portefeuille</b>	2	2	2	1	4	6	4
<b>La gestion des bases de données</b>	3	2	3	2	5	5	7
<b>La gestion des actifs</b>	3	2	5	4	5	4	5
<b>La gestion des arrêts et des pannes (maintenance)</b>	2	2	4	3	4	5	2
<b>Hygiène et sécurité</b>	3	3	5	2	2	2	2

ENAFOR							
	1	2	3	4	5	6	7
Finance	6	6	4	6	7	7	7
Ressources humaines	6	5	4	6	6	7	7
Recherche et développement	1	1	1	1	1	2	1
Ingénierie	3	2	2	3	3	2	2
Logistique	4	4	3	4	5	3	4
La gestion des projets et du portefeuille	4	3	4	4	4	6	4
La gestion des bases de données	3	4	5	4	6	6	7
La gestion des actifs	6	5	4	5	6	4	6
La gestion des arrêts et des pannes (maintenance)	6	5	5	5	5	4	3
Hygiène et sécurité	3	3	4	2	2	2	2
HYPROC							
	1	2	3	4	5	6	7
Finance	6	4	4	4	7	4	4
Ressources humaines	5	5	4	3	6	4	5
Recherche et développement	1	1	1	1	1	2	1
Ingénierie	1	2	2	3	3	2	2
Logistique	4	4	3	3	5	3	4
La gestion des projets et du portefeuille	4	3	2	1	4	6	4
La gestion des bases de données	3	2	3	3	6	5	7
La gestion des actifs	6	5	4	4	6	4	5
La gestion des arrêts et des pannes (maintenance)	5	5	5	3	4	6	3
Hygiène et sécurité	3	3	4	2	2	2	2

Tableau 2: données sur la maturité des SI de NAFTAL, ENAFOR et HYPROC

### 3.1.3 l'analyse des résultats

Cette partie s'intéresse aux résultats obtenus par l'application de l'algorithme du clustering sur les deux types de données, compte tenu de la présence de deux clustering, cela implique la nécessité de considérer deux analyses : une première sur les données des filiales et une autre sur la maturité des SI par filiale.

#### 3.1.3.1 l'analyse du clustering des données sur les filiales

Les résultats de l'algorithme sont les suivants :

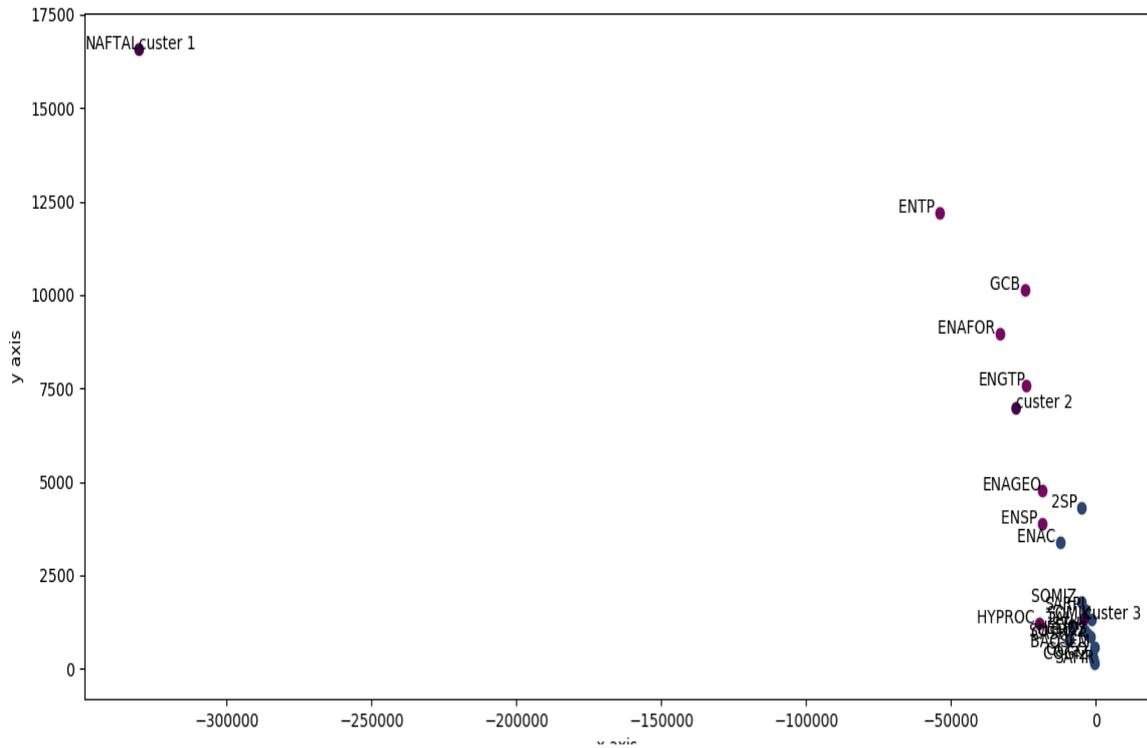


figure 17: résultat du clustering des filiales

```

----- Clustering Report -----

-----cluster:1-----
NAFTAL

-----cluster:2-----
ENAGEO
ENTP
ENSP
ENAFOR
ENGTP
GCB
HYPROC

-----cluster:3-----
CASH
SARPI
SOMIK
SOMIZ
SOTRAZ
GCCO
2SP
STH
BAOSEM
SAFIR
ENAC
COGIZ
TAL
HELIOS

-----centroids-----
Cent1 : [332739.0, 8193.0, 30709.0, 37.0]
Cent2 : [27426.428571428572, 4608.428571428572, 6520.142857142857, 36.42857142857143]
Cent3 : [3975.214285714286, 596.9142857142857, 1452.142857142857, 21.785714285714285]
    
```

figure 18: rapport sur le clustering des filiales

**Il ressort de l'application de l'outil que l'on peut regrouper les filiales en 3 clusters. Le cluster Top prioritaire ne contient qu'une seule filiale à savoir Naftal qui devra impérativement être intégrée dans la feuille de route. Le cluster de 2<sup>nd</sup>e priorité regroupe 7 filiales importantes. Enfin, le cluster 3 regroupe les filiales de moindre priorité.**

**Par la suite nous retiendrons dans l'étude Naftal (Top prioritaire) ainsi que Hyproc et Enafor en raison de leur niveau de priorité (Cluster 2) et de la possibilité d'avoir un accès aisé à leur management.**

**Des explications détaillées seront fournies dans l'analyse ci-dessous.**

### L'analyse

L'application de l'algorithme des k-moyennes sur les données des filiales a produit les résultats suivants, tels que les montrent les images précédentes (figure 19 et 20) :

Le premier cluster inclut NAFTAL, avec le centroïde (332739.0, 8193.0, 30709.0, 37.0), ce dernier inclut les valeurs maximales de tous les types de données, chiffre d'affaire, résultats, personnel et expérience, et cela indique l'importance de cette filiale dans la valeur ajoutée globale du groupe,

Le deuxième cluster inclut ENAGEO, ENTP, ENSP, ENAFOR, ENGTP, GCB et HYPROC avec le centroïde (27426.428571428572, 4608.428571428572, 6520.142857142857, 6.42857142857143), qui affiche des valeurs qui sont au-dessus de la moyenne des données, cela montre que ces filiales ne sont pas aussi importantes que NAFTAL, mais elles doivent être considérées et prises en compte en seconde position après NAFTAL.

Le troisième cluster inclut tout le reste des filiales, avec le centroïde (3975.214285714286, 596.9142857142857, 1452.142857142857, 21.785714285714285), qui montre des valeurs relativement faibles, en d'autres termes ces filiales sont à faible valeur ajoutée comparées à celles des deux premiers clusters, cela rend légitime de retarder la mise en place des solutions pour ce groupe de filiales.

#### 3.1.3.2 l'analyse du clustering des données sur les SI des filiales

- **NAFTAL**

Les résultats de l'algorithme sont les suivants :

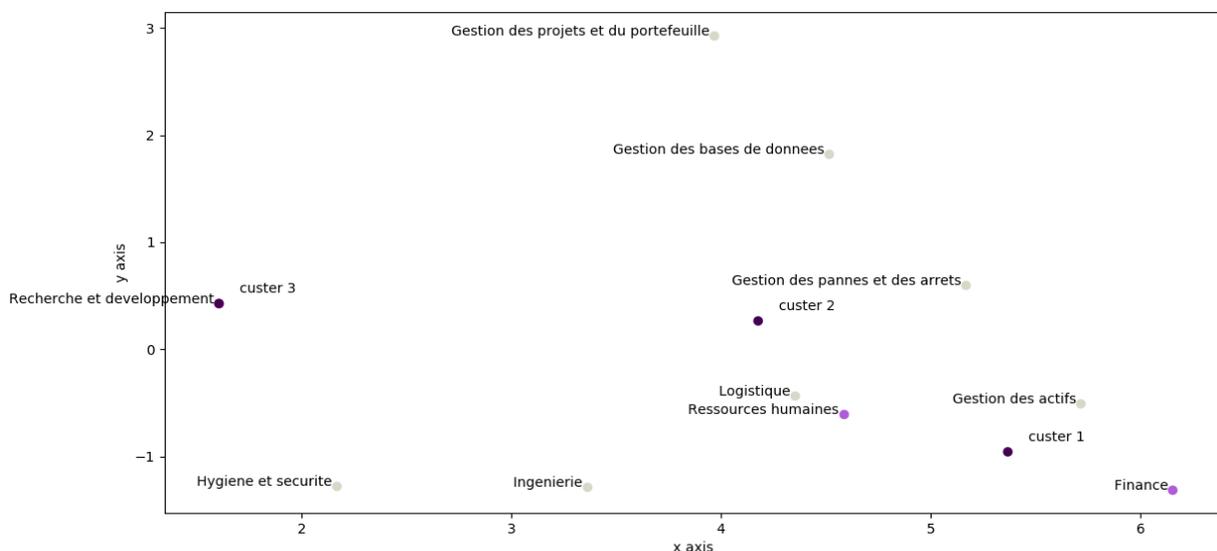


figure 19: résultat du clustering pour NAFTAL

```
----- Clustering Report -----  
  
-----cluster:1-----  
Finance  
Ressources humaines  
  
-----cluster:2-----  
Ingenierie  
Logistique  
Gestion des projets et du portefeuille  
Gestion des bases de donnees  
Gestion des actifs  
Gestion des pannes et des arrêts  
Hygiene et securite  
  
-----cluster:3-----  
Recherche et developpement  
-----centroids-----  
Cent1 : [4.0, 3.5, 5.5, 3.5, 6.0, 5.0, 2.0]  
Cent2 : [2.5714285714285716, 2.2857142857142856, 3.2857142857142856, 2.5714285714285716, 4.0, 4.0, 3.5714285714285716]  
Cent3 : [1.0, 1.0, 1.0, 1.0, 1.0, 2.0, 1.0]
```

*figure 180: rapport du clustering de NAFTAL*

### L'analyse

On constate la formation de trois clusters :

Le premier comporte « Finance » et « ressources humaines », avec le centroïde (4.0, 3.5, 5.5, 3.5, 6.0, 5.0, 2.0), qui comporte les plus grandes valeurs, comparé aux autres clusters, cela implique que ce cluster regroupe les solutions dont la maturité est la plus grande.

Le deuxième cluster regroupe « ingénierie », « logistique », « gestion des projets et du portefeuille », « gestion des bases de données », « gestion des actifs », « gestion des pannes et des arrêts », « hygiène et sécurité », ces dernières sont de maturité relativement faible, tel que le montre leur centroïde (2.57,2.28,3.28,2.57,4,4,3.57), parmi ces systèmes, il existe ceux qui sont fortement liés à l'activité de la filiale, telle que la « logistique », d'où l'importance significative de ce cluster.

Le dernier cluster ne compte qu'une seule solution qui est « recherche et développement », avec un centroïde de très faibles valeurs (1,1,1,1,1,2,1), ce qui implique la faible maturité, voire l'absence de cette activité au sein de cette filiale.

- ENAFOR

Les résultats de l’algorithme sont les suivants :

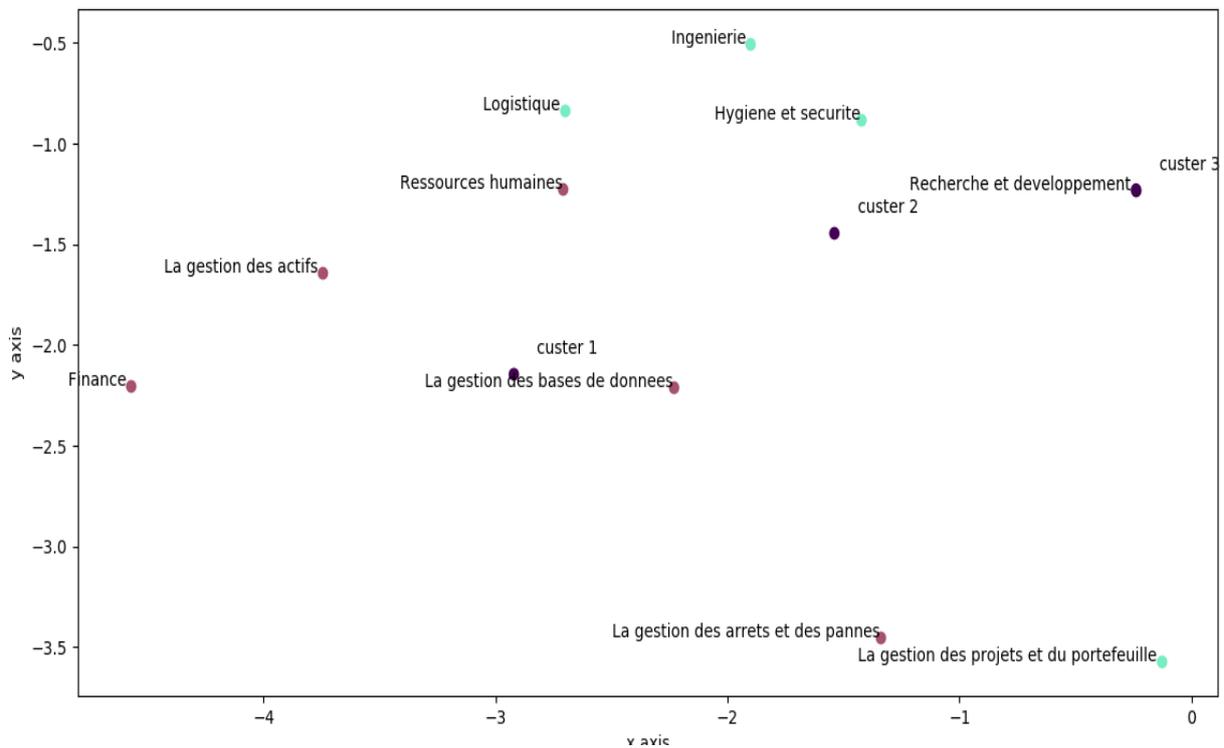


figure 21: résultat du clustering pour ENAFOR

```

----- Clustering Report -----

-----cluster:1-----
Finance
Ressources humaines
La gestion des bases de donnees
La gestion des actifs
La gestion des arrêts et des pannes

-----cluster:2-----
Ingenierie
Logistique
La gestion des projets et du portefeuille
Hygiene et securite

-----cluster:3-----
Recherche et developpement

-----centroids-----
Cent1 : [5.0, 4.2, 4.0, 3.4, 5.8, 4.6, 4.8]
Cent2 : [3.0, 3.0, 2.75, 2.25, 3.5, 3.25, 3.0]
Cent3 : [1.0, 1.0, 1.0, 1.0, 1.0, 2.0, 1.0]
    
```

figure 22: rapport du clustering d'ENAFOR

### L'analyse

La distribution des solutions dans les différents clusters est la suivante :

Le premier cluster inclut « finance », « la gestion des bases de données », « la gestion des actifs », « la gestion des arrêts et des pannes », compte tenu des valeurs élevées de leur centroïde (5.0,4.2,4.0,3.4,5.8,4.6,4.8), on déduit qu'elles sont matures.

Les solutions du deuxième cluster « ingénierie », « logistique », « la gestion des projets et du portefeuille », « hygiène et sécurité » sont d'une importance vitale pour le cœur de l'activité de la filiale, mais elles sont de maturité moyennement faible, cela en se basant sur les valeurs de leur centroïde (3,3,2.75,2.25,3.5,3.25,3).

Le dernier cluster inclut « recherche et développement » avec une très faible maturité, selon son centroïde (1,1,1,1,1,2,1).

- **HYPROC**

Les résultats de l'algorithme sont les suivants :

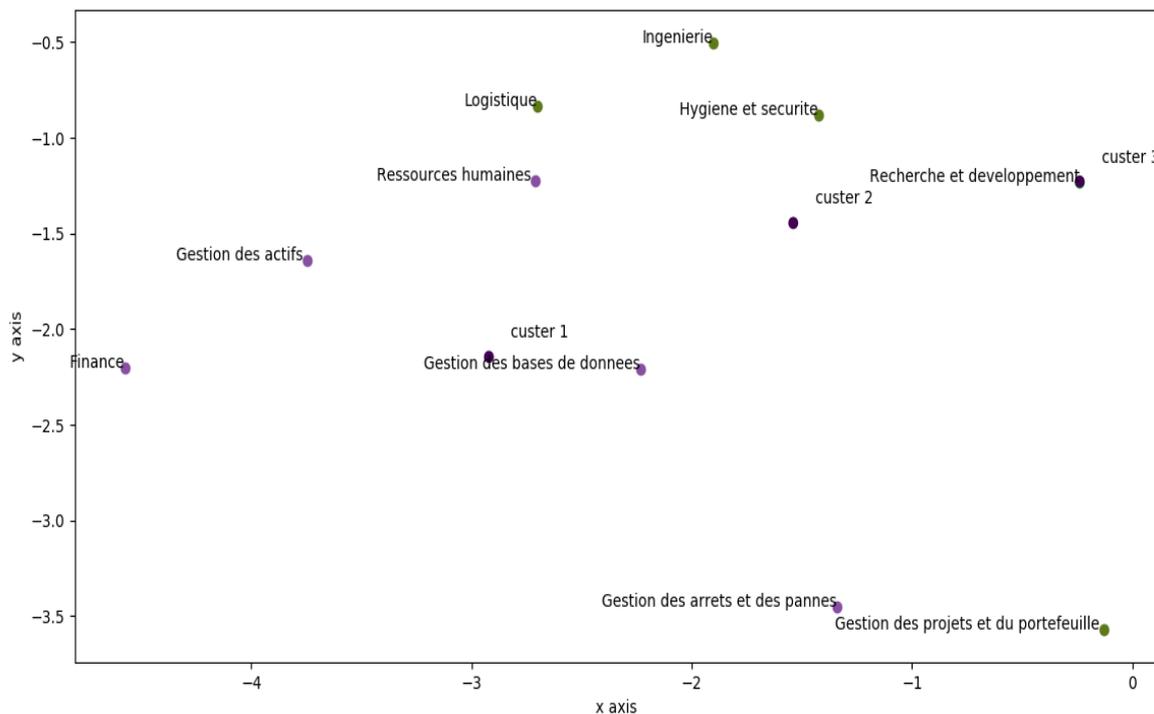


figure 23: résultat du clustering pour HYPROC

```
----- Clustering Report -----  
  
-----cluster:1-----  
Finance  
Ressources humaines  
Gestion des bases de donnees  
Gestion des actifs  
Gestion des arrêts et des pannes  
  
-----cluster:2-----  
Ingenierie  
Logistique  
Gestion des projets et du portefeuille  
Hygiene et securite  
  
-----cluster:3-----  
Recherche et developpement  
-----centroids-----  
Cent1 : [5.0, 4.2, 4.0, 3.4, 5.8, 4.6, 4.8]  
Cent2 : [3.0, 3.0, 2.75, 2.25, 3.5, 3.25, 3.0]  
Cent3 : [1.0, 1.0, 1.0, 1.0, 1.0, 2.0, 1.0]
```

*figure 24: rapport du clustering d'HYPROC*

### L'analyse

Les clusters des solutions de cette filiale sont :

« Finance », « ressources humaines », « gestion des bases de données », « gestion des actifs », « gestion des arrêts et des pannes » représentent les composantes du premier cluster, avec un niveau de maturité appréciable, cela étant justifié par les valeurs du centroïde (5,4.2,4,3.4,5.8,4.6,4.8).

Le deuxième cluster inclut majoritairement des solutions, dont la performance de l'activité en dépend fortement, ces solutions sont « ingénierie », « logistique », « gestion des projets et du portefeuille » et « hygiène et sécurité » ; leur centroïde (3,3,2.75,2.25,3.5,3.25,3) témoigne d'une faible maturité

A l'instar des filiales précédentes, « recherche et développement » forme le dernier cluster avec un centroïde de très faibles valeurs (1,1,1,1,1,2,1), d'où une très faible maturité.

### 3.2 L'élaboration de la feuille de route

On constate dans les trois analyses des différentes filiales, l'existence d'une similitude dans la distribution des filiales sur les différents clusters. Cette similitude peut être illustrée comme suit :

Le premier cluster n'inclut que les solutions de gestion courante, avec une maturité élevée, ensuite concernant les deuxièmes clusters, ils incluent généralement les solutions qui sont directement liées au cœur du métier de la filiale. Cela s'interprète par la grande importance de ce cluster, mais en revanche sa maturité est faible, et pour le dernier cluster, il inclut « recherche et développement » avec une très faible maturité, cela s'explique par le fait que ce département a été récemment mis en place au sein du groupe.

En considérant les résultats et les conclusions tirés des différentes analyses, la feuille de route sera réalisée en suivant cette approche :

En premier lieu, les filiales seront priorisées selon les clusters à grandes valeurs, car ces derniers sont associés à une grande valeur ajoutée, ensuite pour chaque filiale, l'implémentation des solutions SAP va suivre une priorisation, qui est basée sur les résultats des différents clusterings. En d'autres termes, les solutions des deuxièmes clusters seront priorisées, car elles sont directement liées à l'activité de la filiale concernée, mais aussi pour pallier à la faible maturité. En deuxième position, viennent les solutions du premier cluster, qui touchent la gestion courante et qui doivent perfectionner la maturité des SI, et le dernier cluster est celui de la « recherche et développement », cette solution sera envisageable une fois que le département R&D sera mis en place et que cette activité prendra de l'ampleur.

Ci-dessous un schéma illustrant la feuille de route :

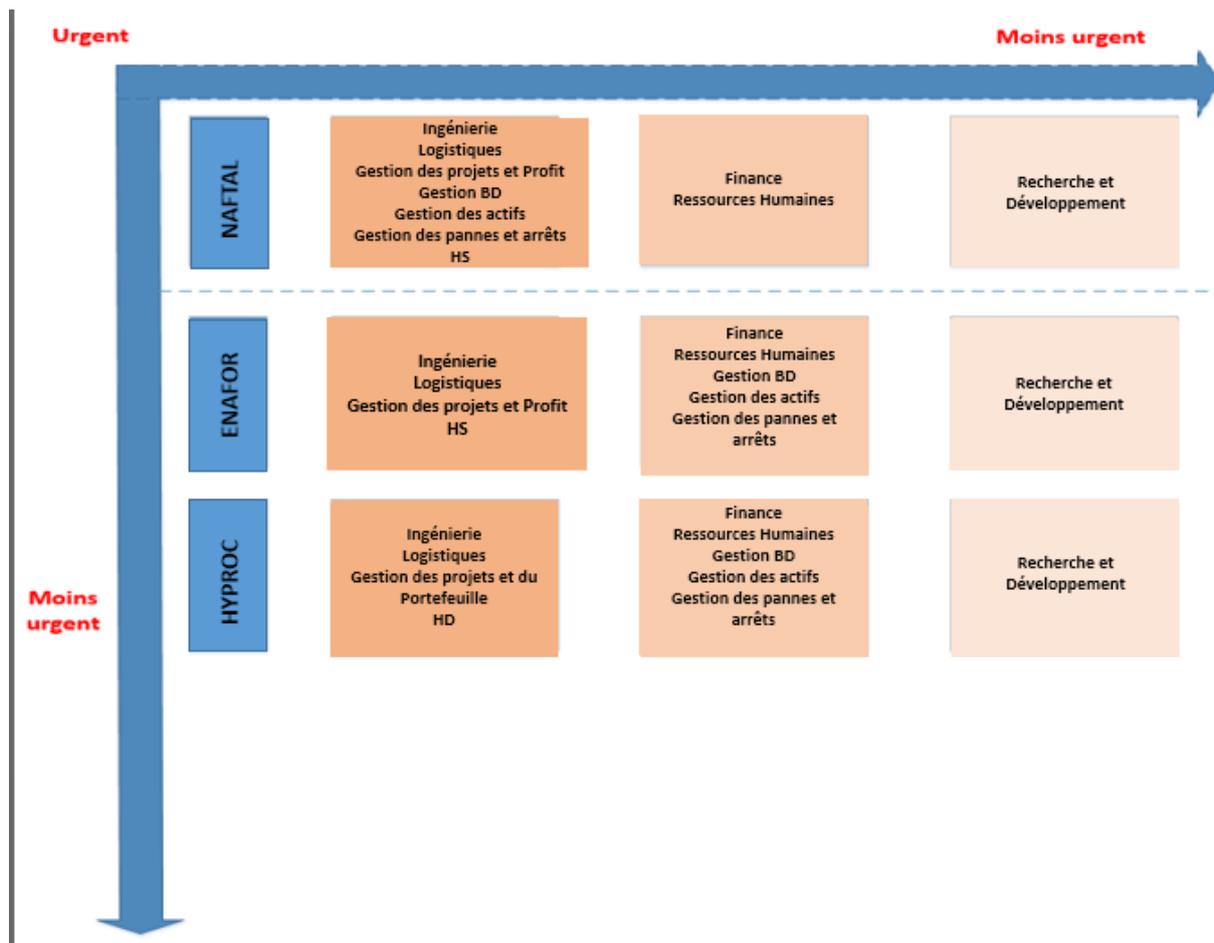


figure 25: schéma explicatif de la feuille de route

### 3.3 Conclusion

Ce chapitre a servi de synthèse au développement de la solution adoptée pour résoudre la problématique du client de SAP, cela en partant de la conception et du développement du logiciel, ensuite la récolte des données, pour arriver au traitement et à l'analyse des résultats.

Ce chapitre s'est achevé par une recommandation illustrée sous forme de feuille de route, cette dernière va orienter le groupe dans sa stratégie de mise en place des solutions SAP.

Les rapports générés par l'outil d'aide à la décision peuvent servir de support pour les éventuels partenaires que le groupe va solliciter pour l'intégration. L'application développée a fait preuve d'un outil efficace, qui a pu résoudre un problème de classification dans le cadre de l'aide à la décision.

# Conclusion générale

---

Un système d'information doté de solutions digitales performantes et intégrées, et qui sont exploitées d'une manière optimale, représente l'un des plus grands challenges et défis de l'entreprise d'aujourd'hui. SONATRACH compte parmi ces entreprises qui envisagent d'avoir une telle structure moderne.

SONATRACH veut avoir une orientation sur le déploiement du portefeuille dont il a fait l'acquisition, et SAP lui assure cette mission de conseil. C'est dans cette optique d'accompagnement que mon travail était organisé, et à son issue, une feuille de route a été élaborée et qui prend en considération les besoins et les impératifs business.

Le cœur de la problématique consistait à définir un programme de déploiement de solutions SAP, avec une vision bidimensionnelle, qui prend en considération les filiales du groupe ainsi que les solutions du portefeuille. Le groupe SONATRACH souhaite avoir un plan qui montre un certain ordre de priorité pour la mise en place de chaque solution, pour chacune des filiales, qui sont préalablement classées et regroupées en fonction de leur importance.

Après avoir analysé et compris ce besoin et ces attentes de SONATRACH, il était possible de les formaliser sous forme d'un problème d'aide à la décision multicritère, de type classification.

L'idée d'utiliser l'apprentissage machine pour la résolution du problème, était née suite à une recherche bibliographique sur les différentes approches de résolution des problèmes d'aide à la décision, et le choix de l'intelligence artificielle se justifie par toutes les qualités et les performances qu'offre cette technique. Ensuite, il y a le passage au développement d'un programme informatique qui assure le traitement des données par l'algorithme d'apprentissage non supervisé, cela en choisissant le langage de programmation « Python » qui est une référence dans le domaine de « data science »

Une fois que le programme informatique a été achevé, il était nécessaire de collecter les données, qui étaient de deux types : le premier concerne et comprend des critères sur les filiales, et le deuxième type reflète la maturité des systèmes d'information et des solutions actuellement utilisées par quelques filiales, choisies pour illustrer la solution adoptée.

L'outil a été développé et les données collectées. Il a été nécessaire d'effectuer le traitement et l'analyse. Une fois ces derniers ont été effectués, des rapports et des graphes ont été générés, et ont servi par la suite comme supports et instruments d'aide à la décision. Enfin, l'aboutissement de toutes ces opérations est la feuille de route.

Ce travail peut être sujet à d'éventuelles améliorations qui peuvent être les suivantes :

- Inclure d'autres méthodes de classifications avec d'autres algorithmes
- Rendre l'outil capable de résoudre d'autres problématiques d'aide à la décision (tri, description ...)
- Trouver d'autres données encore plus pertinentes pour juger les filiales.
- Une meilleure interface graphique.

# **Bibliographie**

---

### Bibliographie

EPINASSE, B., 2001. *Les systèmes d'information opérationnels de l'entreprise*. [En ligne] Disponible à l'adresse: <http://www.lsis.org/epinasseb/Supports/SIO/4-SIOperationnels-4p.pdf>

HENRIET, L., 2000. *Systèmes d'évaluation et de classification multicritères pour l'aide à la décision : construction de modèles et procédures d'affectation*, Paris: Université Paris Dauphine.

IBM Software Group, 2007. *introduction to business intelligence*, s.l.: IBM Corporation.

MAHARRAR, A., 2014. *la mise en place d'un système d'information formalisé dans les entreprises algériennes*, Tlemcen: Université de Tlemcen.

MAMMERY, M., 2013. *Une approche d'aide multicritère à la décision pour l'évaluation du confort dans les trains : construction d'un modèle d'évaluation*, Paris: Université Paris Dauphine.

O'BRIEN, J., 2003. *introduction aux systèmes d'information*. 2nd éd. s.l.:s.n.

ROY, B., 1985. *Méthodologie multicritère d'aide à la décision*. s.l.:s.n.

RUSSEL, S., 1994. *Artificial Intelligence, a modern approach*. 3rd éd. s.l.:Prentice Hall.

ULLMAN, S., 2014. *Unsupervised learning : Clustering*. s.l.:Center for Brains, Minds & Machine.

WILD, V. & ROMAGNI, P., 1998. *l'intelligence économique au service de l'entreprise*. Paris: LES PRESSES DU MANAGEMENT.

### Webographie

**Web a** : [http://www.assises.club-mes.com/MES\\_manufacturing\\_execution\\_system.asp#titre13](http://www.assises.club-mes.com/MES_manufacturing_execution_system.asp#titre13)

**Web b** : <http://www.institut-numerique.org/1-definition-de-linformation-5028f58b86944>

**Web c** : <https://www.afis.fr/nm-is/Pages/Ing%C3%A9nierie%20Syst%C3%A8me/Le%20syst%C3%A8me%20et%20sa%20d%C3%A9finition.aspx>

# **Annexes**

---

## Annexe 1 : Le mode d'emploi du logiciel de clustering

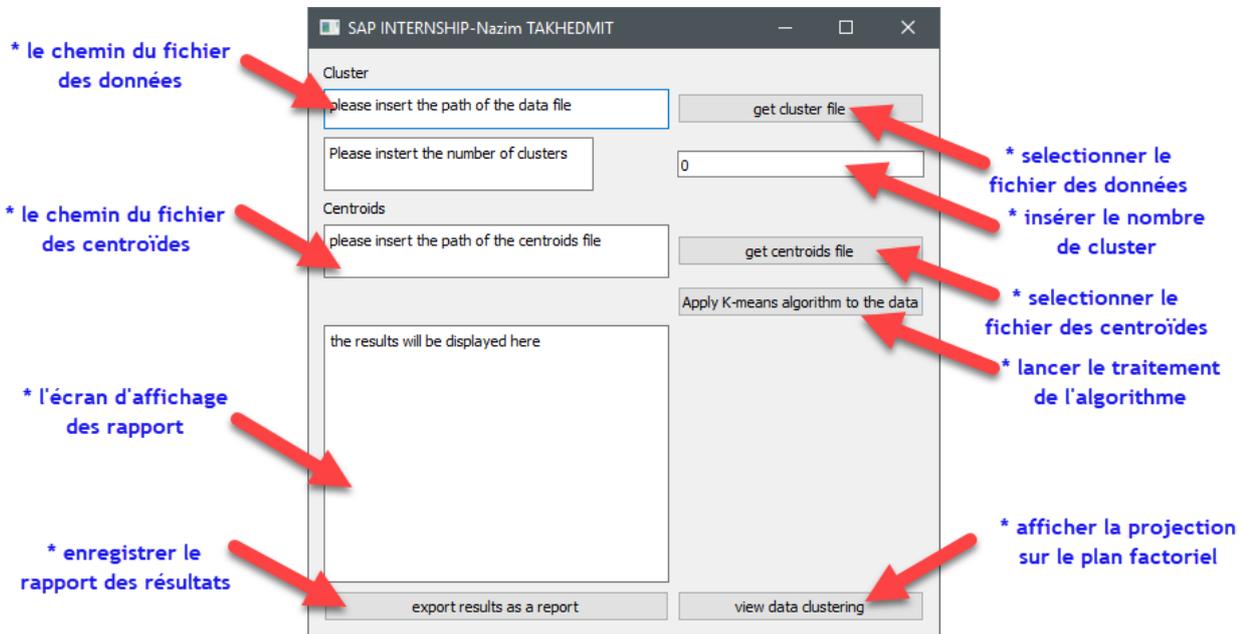


figure 19: mode d'emploi du programme informatique

## Annexe 2 : Le code source du logiciel

```
from PyQt5.QtWidgets import *
from PyQt5.QtCore import *
import os
import sys
import csv
import array
import math
import numpy as np
from numpy import linalg as LA
import matplotlib.pyplot as plt
import random

class Window(QWidget):
    def __init__(self):
        QWidget.__init__(self)
        self.setWindowTitle("SAP INTERNSHIP-Nazim TAKHEDMIT")

        self.layout = QGridLayout()
        self.setLayout(self.layout)

        self.label = QLabel("Cluster")
        self.layout.addWidget(self.label, 0, 0)

        self.textbr1 = QTextBrowser()
        self.textbr1.insertPlainText("please insert the path of the data file")
```

```

self.textbr1.setMaximumSize(400,30)
self.layout.addWidget(self.textbr1, 1, 0)

self.btn1=QPushButton("get cluster file")
self.layout.addWidget(self.btn1, 1, 1)
self.btn1.clicked.connect(self.getCtrFilepath)

self.clsnum = QLineEdit("Cluster's number")
self.clsnum.setText("0")
self.layout.addWidget(self.clsnum, 2, 1)

self.dataClst = int(0)
self.dataCent = int(0)

self.textbr2 = QTextBrowser()
self.textbr2.insertPlainText("Please instert the number of clusters")
self.textbr2.setMaximumSize(200, 40)
self.layout.addWidget(self.textbr2, 2, 0)

self.label3 = QLabel("Centroids")
self.layout.addWidget(self.label3, 3, 0)

self.textbr3 = QTextBrowser()
self.textbr3.insertPlainText("please insert the path of the centroids file")
self.textbr3.setMaximumSize(400, 40)
self.layout.addWidget(self.textbr3, 4, 0)

self.btn2 = QPushButton("get centroids file")
self.layout.addWidget(self.btn2, 4, 1)
self.btn2.clicked.connect(self.getCtdFilepath)

self.msg = QMessageBox()
self.msg.setText("Pease check your inputs !")
self.msg.setInformativeText("clusters number is different from the centroids matrix lenght")
self.msg.setIcon(QMessageBox.Information)
self.msg.setStandardButtons(QMessageBox.Close)

self.msg2 = QMessageBox()
self.msg2.setText("Pease check your inputs !")
self.msg2.setInformativeText("insufficient inputs !")
self.msg2.setIcon(QMessageBox.Information)
self.msg2.setStandardButtons(QMessageBox.Close)

self.btnStart = QPushButton("Apply K-means algorithm to the data")
self.layout.addWidget(self.btnStart, 5, 1)
self.btnStart.clicked.connect(self.brain)

self.textbr4 = QTextBrowser()
self.textbr4.insertPlainText("the results will be displayed here")
self.textbr4.setMaximumSize(700, 200)
self.layout.addWidget(self.textbr4, 6, 0)

```

```

self.btnview = QPushButton("view data clustering")
self.layout.addWidget(self.btnview,7, 1)
self.btnview.clicked.connect(self.viewd)

self.export = QPushButton("export results as a report")
self.layout.addWidget(self.export, 7, 0)
self.export.clicked.connect(self.exportd)

self.text="\n----- Clustering Report -----"

def showError(self):
    self.layout.removeWidget(self.textbr4)
    textbr4 = QTextBrowser()
    textbr4.setMaximumSize(700, 200)
    textbr4.setPlainText("please consider the error message, then review the inputs")
    self.layout.addWidget(textbr4, 5, 0)
    d2 = open(self.path2[0], 'r')
    datacsv2 = csv.reader(d2)
    centroids = []
    for i in datacsv2:
        l = []
        l = [float(k) for k in i]
        centroids.append(l)

    if(int(self.clsnum.text())!=len(centroids)):
        self.msg.show()
        return 0
    else:
        return 1

def showError2(self):
    self.layout.removeWidget(self.textbr4)
    textbr4 = QTextBrowser()
    textbr4.setMaximumSize(700, 200)
    textbr4.setPlainText("please consider the error message, then review the inputs")
    self.layout.addWidget(textbr4, 5, 0)
    if (self.dataCent==0 and self.dataClst == 0):
        self.msg2.show()
        return 0
    else:
        return 1

def getCtrFilepath(self):
    self.layout.removeWidget(self.textbr1)
    self.path1 = QFileDialog.getOpenFileName(self, caption='Get a path from', filter='All
Files(*.*)')
    textbr1 = QTextBrowser()
    textbr1.setMaximumSize(400, 40)
    textbr1.setPlainText(self.path1[0])
    self.layout.addWidget(textbr1, 1, 0)
    self.dataClst = self.dataClst +1

def exportd(self):

```

```

name = QFileDialog.getSaveFileName(self, 'Save File')
txt_file=open(name[0], 'w')
for e in self.strD:
    txt_file.write(str(e))
for c in self.clname:
    txt_file.write(str(c))
txt_file.close()

```

```

def getCtdFilepath(self):
    self.layout.removeWidget(self.textbr3)
    self.path2 = QFileDialog.getOpenFileName(self, caption='Get a path from', filter='All
Files(*.*)')
    textbr3 = QTextBrowser()
    textbr3.setMaximumSize(400, 40)
    textbr3.setPlainText(self.path2[0])
    self.layout.addWidget(textbr3, 4, 0)
    self.dataCent = self.dataCent+1

```

```

def brain(self):
    if(self.showError2() != 0):
        if(self.showError() != 0):
            self.layout.removeWidget(self.textbr4)
            d = open(self.path1[0], 'r')
            datacsv = csv.reader(d)
            matrix = []
            x = []
            for i in datacsv:
                l = []
                x.append(i[0])

                for k in range(1, len(i)):
                    l.append(float(i[k]))
                matrix.append(l)

            dict = {}
            for i in range(0, len(x)):
                dict[x[i]] = matrix[i]

            a = []
            a.append(dict)
            a.append(matrix)

            d2 = open(self.path2[0], 'r')
            datacsv2 = csv.reader(d2)
            centroids = []
            for i in datacsv2:
                l = []
                l = [float(k) for k in i]
                centroids.append(l)

            def dist(x, y):

```

```

if len(x) != len(y):
    print("data variables have different lengths")
else:
    d = 0;
    i = 0;
    while i < len(y):
        d = d + pow((x[i] - y[i]), 2)
        i += 1
    return math.sqrt(d)

def avg(l):
    s = 0
    for i in l:
        s = s + i
    return s / len(l)

def cnew(m):
    Cn = []

    for u in m:
        ut = np.transpose(u)
        ct = []
        for t in ut:
            ct.append(avg(t))
        Cn.append(ct)
    return Cn
clstr=[]

def km(a, b):
    cn = []
    while (b != cn):
        C = []

        for i in range(0, len(b)):
            C.append([])

        d = []
        for k in a:
            for m in b:
                d.append(dist(k, m))

        C[d.index(min(d))].append(k)
        d = []
        cn = b
        b = cnew(C)
        self.clstr =b

    return C

c = km(a[1], centroids)

self.clname = []

```

```

hn = 1
for k in self.clstr:
    if (hn == 1):
        self.cname.append("\r\n"+"-----centroids-----")

        self.cname.append("\nCent" + str(hn) + " : " + str(k))
        hn = hn + 1

print(self.cname)

i = 1
s = ""
self.strD = []
self.strD.append(self.text)
for e in c:
    s = s + "\n-----cluster:" + str(i) + "-----"
    self.strD.append("\r\n"+"n-----cluster:" + str(i) + "-----")
    for k in e:
        for key, val in a[0].items():
            if k == val:
                s = s + "\n" + key
                self.strD.append("\n" + str(key))
    i = i + 1

textbr4 = QTextBrowser()
textbr4.setMaximumSize(700, 200)
textbr4.setPlainText(s)
self.layout.addWidget(textbr4, 5, 0)

def pl(a, b):
    s = 0
    if (len(np.transpose(a)) != len(np.transpose(b))):
        print("please check your matrix length")
    else:
        for i in range(0, len(a)):
            s = s + a[i] * b[i]
    return s

def pm(a, b):
    if (len(np.transpose(a)) != len(np.transpose(b))):
        print("please check your matrix length")
    else:
        C = []
        for x in a:
            D = []
            for y in b:
                D.append(pl(x, y))
            C.append(D)
            D = []
        return C

res = pm(np.transpose(matrix), np.transpose(matrix))
Egm = LA.eig(res)
vect1 = []

```

```

vect1.append(list(Egm[1][list(Egm[0]).index(max(list(Egm[0])))]))
e = list(Egm[0])
e.remove(max(e))
vect1.append(list(Egm[1][list(Egm[0]).index(max(e))]))

d=dict
for k,e in d.items():
    h=pm([e],vect1)
    d[k]=h[0]

for k in c:
    new = np.transpose(pm(k, vect1))

plt.scatter(new[0], new[1],
            c=(random.randint(1, 100) / 100, random.randint(1, 100) / 100,
              random.randint(1, 100) / 100))

lab = []
for a in np.transpose(new):

    for key, val in d.items():

        if list(a) == val:
            lab.append(key)

for i in range(0, len(lab)):
    plt.annotate(lab[i], [new[0][i], new[1][i]],ha='right' )

dcstr = { }
for i in range(0, len(self.clstr)):
    dcstr['custer %d' %(i+1)] = self.clstr[i]

for k,e in dcstr.items():
    h=pm([e],vect1)
    dcstr[k]=h[0]

print(dcstr)

n=[]
k=[]
for a,e in dcstr.items():
    n.append(e)
    k.append(a)
new=np.transpose(n)
plt.scatter(new[0], new[1],c=(1,1,1),)
for i in range(0, len(k)):
    plt.annotate(k[i], [new[0][i], new[1][i]],ha='left',xytext=(new[0][i]+0.1, new[1][i]+0.1))

```

```
plt.xlabel('x axis')  
plt.ylabel('y axis')
```

```
def viewd(self):
```

```
    plt.show()
```

```
app = QApplication(sys.argv)  
screen = Window()  
screen.show()  
sys.exit(app.exec_())
```

### Annexe 3 : Le formulaire des filiales

Est-ce que vous disposez de solutions informatiques qui couvrent les activités suivantes: \*

	Oui	Non
Finance	<input type="checkbox"/>	<input type="checkbox"/>
RH	<input type="checkbox"/>	<input type="checkbox"/>
Recherche et développement	<input type="checkbox"/>	<input type="checkbox"/>
Ingénierie	<input type="checkbox"/>	<input type="checkbox"/>
logistique	<input type="checkbox"/>	<input type="checkbox"/>
gestion des projets et du portefeuille	<input type="checkbox"/>	<input type="checkbox"/>
La gestion des bases de données	<input type="checkbox"/>	<input type="checkbox"/>
La gestion des actifs	<input type="checkbox"/>	<input type="checkbox"/>
gestion des arrêts et des pannes de production	<input type="checkbox"/>	<input type="checkbox"/>
Hygiène et sécurité	<input type="checkbox"/>	<input type="checkbox"/>

Est-ce que les résultats et les décisions que proposent vos solutions sont exploitables ? \*

1 sur 8    2 sur 8    3 sur 8    4 sur 8    5 sur 8    6 sur 8    7 sur 8    8 sur 8

Finance	<input type="checkbox"/>							
RH	<input type="checkbox"/>							
Recherche et développement	<input type="checkbox"/>							
Ingénierie	<input type="checkbox"/>							
logistique	<input type="checkbox"/>							
gestion des projets et du portefeuille	<input type="checkbox"/>							
La gestion des bases de données	<input type="checkbox"/>							
La gestion des actifs	<input type="checkbox"/>							
gestion des arrêts et des pannes de production	<input type="checkbox"/>							
Hygiène et sécurité	<input type="checkbox"/>							

Est-ce que vous exploitez toutes les fonctionnalités de vos solutions informatiques ? \*

	1 sur 8	2 sur 8	3 sur 8	4 sur 8	5 sur 8	6 sur 8	7 sur 8	8 sur 8
Finance	<input type="checkbox"/>							
RH	<input type="checkbox"/>							
Recherche et développement	<input type="checkbox"/>							
Ingénierie	<input type="checkbox"/>							
logistique	<input type="checkbox"/>							
gestion des projets et du portefeuille	<input type="checkbox"/>							
La gestion des bases de données	<input type="checkbox"/>							
La gestion des actifs	<input type="checkbox"/>							
gestion des arrêts et des pannes de production	<input type="checkbox"/>							
Hygiène et sécurité	<input type="checkbox"/>							

Est-ce que vos solutions couvrent l'intégralité de vos besoins dans les activités suivantes ? \*

	1 sur 8	2 sur 8	3 sur 8	4 sur 8	5 sur 8	6 sur 8	7 sur 8	8 sur 8
Finance	<input type="checkbox"/>							
RH	<input type="checkbox"/>							
Recherche et développement	<input type="checkbox"/>							
Ingénierie	<input type="checkbox"/>							
logistique	<input type="checkbox"/>							
gestion des projets et du portefeuille	<input type="checkbox"/>							
La gestion des bases de données	<input type="checkbox"/>							
La gestion des actifs	<input type="checkbox"/>							
gestion des arrêts et des pannes de production	<input type="checkbox"/>							
Hygiène et sécurité	<input type="checkbox"/>							

## Est-ce que vous jugez que les solutions vous aident concrètement? \*

	1 sur 8	2 sur 8	3 sur 8	4 sur 8	5 sur 8	6 sur 8	7 sur 8	8 sur 8
Finance	<input type="checkbox"/>							
RH	<input type="checkbox"/>							
Recherche et développement	<input type="checkbox"/>							
Ingénierie	<input type="checkbox"/>							
logistique	<input type="checkbox"/>							
gestion des projets et du portefeuille	<input type="checkbox"/>							
La gestion des bases de données	<input type="checkbox"/>							
La gestion des actifs	<input type="checkbox"/>							
gestion des arrêts et des pannes de production	<input type="checkbox"/>							
Hygiène et sécurité	<input type="checkbox"/>							

## Est-ce que le personnel est à l'aise avec ces solutions ? \*

	1 sur 8	2 sur 8	3 sur 8	4 sur 8	5 sur 8	6 sur 8	7 sur 8	8 sur 8
Finance	<input type="checkbox"/>							
RH	<input type="checkbox"/>							
Recherche et développement	<input type="checkbox"/>							
Ingénierie	<input type="checkbox"/>							
logistique	<input type="checkbox"/>							
gestion des projets et du portefeuille	<input type="checkbox"/>							
La gestion des bases de données	<input type="checkbox"/>							
La gestion des actifs	<input type="checkbox"/>							
gestion des arrêts et des pannes de production	<input type="checkbox"/>							
Hygiène et sécurité	<input type="checkbox"/>							

veuillez attribuer un score pour le degré d'informatisation(indépendance par rapport au support papier) des départements suivants \*

	1 sur 8	2 sur 8	3 sur 8	4 sur 8	5 sur 8	6 sur 8	7 sur 8	8 sur 8
Finance	<input type="checkbox"/>							
RH	<input type="checkbox"/>							
Recherche et développement	<input type="checkbox"/>							
Ingénierie	<input type="checkbox"/>							
logistique	<input type="checkbox"/>							
gestion des projets et du portefeuille	<input type="checkbox"/>							
La gestion des bases de données	<input type="checkbox"/>							
La gestion des actifs	<input type="checkbox"/>							
gestion des arrêts et des pannes de production	<input type="checkbox"/>							
Hygiène et sécurité	<input type="checkbox"/>							

