



Département Génie Industriel

Mémoire de projet de fin d'études

Pour l'obtention du diplôme d'Ingénieur d'Etat en Génie Industriel

**Contribution à l'optimisation du processus de gestion des conteneurs par
l'apprentissage non-supervisé**

Cas des conteneurs en souffrance de MAERSK Algérie

Présenté par :

Kadi Mohamed et Bouhali Ghezlane

Sous la direction de :

Mme Sihem FELLOUS (MAERSK)

Mme Bahia BOUCHAFAA (ENP)

M. Reda GOURINE (ENP)

Présenté et soutenu publiquement le (08/07/2021) devant le jury composé de :

Présidente	Mme S. Ait BOUAZZA	MAA	ENP
Promoteur	Mme B. BOUCHAFAA	MAA	ENP
Promoteur	Mr R. GOURINE	Docteur	ENP
Examineur	Mr O. ARKI	MCB	ENP



Département Génie Industriel

Mémoire de projet de fin d'études

Pour l'obtention du diplôme d'Ingénieur d'Etat en Génie Industriel

**Contribution à l'optimisation du processus de gestion des conteneurs par
l'apprentissage non-supervisé**

Cas des conteneurs en souffrance de MAERSK Algérie

Présenté par :

Kadi Mohamed et Bouhali Ghezlane

Sous la direction de :

Mme Sihem FELLOUS (MAERSK)

Mme Bahia BOUCHAFAA (ENP)

M. Reda GOURINE (ENP)

Présenté et soutenu publiquement le (08/07/2021) devant le jury composé de :

Présidente	Mme S. Ait BOUAZZA	MAA	ENP
Promoteur	Mme B. BOUCHAFAA	MAA	ENP
Promoteur	Mr R. GOURINE	Docteur	ENP
Examineur	Mr O. ARKI	MCB	ENP

Dédicace

“

Je dédie ce travail

À ma mère, qui a souffert sans me laisser souffrir, qui n'a épargné aucun effort pour me voir réussir, avec son amour, sa motivation et ses sacrifices. À mon père, pour son amour, son encouragement et son soutien qu'il m'a toujours apporté.

À mon cher grand père Larbi, mon héros, mon plus fort repère et ma fierté. À ma chère grand-mère Hadda, qui m'a apporté tant d'amour et de soutien. Que dieu les protège et leur procure une bonne santé et une longue vie.

À ma chère tante Sabrina et son mari Said, mes chers oncles Nassim et Zohir qui n'ont pas cessé de m'encourager et de me soutenir tout au long de mes études.

À mon cher binôme Mohamed pour tous les efforts qu'il a fourni durant ce projet.

À mes chers amis : Chahinez, Djazia, Nadia, Ramzi, Wadoud, Mahyou, Roro, Belkis, Halla, Yacine, Abdallah, Yassine, Chaouki, Bilel et plus particulièrement à ma chère team BOUCHAHLATA, avec qui j'ai partagé des moments inoubliables.

À tous les membres de IEC/VIC.

”

- *Ghezlane*

Dédicace

“

Je dédie ce travail

*À mes chers parents pour tous leurs sacrifices, leur amour,
leur dévouement et leur soutien tout au long de mes études.*

À mes chers frères Riad & Farid.

*À la mémoire de mon grand-père que Dieu l'accueille dans
son vaste paradis.*

*À mes grands-mères et mon grand-père qui n'ont jamais
cessé de prier pour moi.*

À ma chère binôme Ghezlane.

*À tous mes amis de comité clubawi, Nacim, Malek, Ramzi,
Otba, Rami et Dalil*

*À tous mes amis, Rayan, Sidali, Karim, Oussama,
Moussa, Danyl, Mehdi et Riad.*

À toute l'équipe collo.

À tout les membres du CAP.

”

- Mohamed

Remerciements

Nous tenons à remercier Allah le tout puissant de nous avoir donné la foi, la force, et la patience pour réaliser ce travail.

Nous remercions nos chers parents, nos frères et soeurs, nos familles qui ont rempli nos vies de joie, de motivation et d'amour.

Nous remercions aussi Mme Bouchafaa et Mr Gourine pour nous avoir encadré et soutenu au long de notre projet ainsi qu'aux membres du jury qui ont accepté d'évaluer notre travail.

Nous remercions toute l'équipe CX de Maersk Algérie pour leur accueil et leur aide, nous remercions plus particulièrement Mme. Sihem Fellous, Melle Sabrina Bonatero, Mr Nasreddine Guerbyenne et Mr Koussaila Hadji.

Nous tenons à remercier surtout, nos collègues, ceux que nous avons croisé durant notre parcours d'études, qui nous ont donné le sentiment d'appartenance, qui nous ont offert de l'aide, du soutien, ou juste du sourire.

Enfin, nous remercions toute personne ayant cru en nous, ayant contribué de près ou de loin, avec beaucoup ou avec peu, à la réalisation de ce travail.

يندرج هذا العمل في إطار تحسين عملية إدارة الحاويات المتأخرة في شركة ميرسك الجزائر ، من خلال تنفيذ نموذج يتمثل في تحسين إدارة الزبائن والحاويات مع مراعاة العوامل المختلفة المؤثرة في الاحتفاظ بالحاويات ونتائجها. يتيح الحل المعنى للشركة استجابة أفضل في إدارة الحاويات المتأخرة فضلا عن توفير الوقت عن طريق اتخاذ إجراءات متباينة بشأن كل مجموعة مشكلة.

الكلمات المفتاحية: الحاويات المتأخرة ، مساعدة اتخاذ القرار ، التعلم الذاتي ، لوحات القيادة

Abstract :

This work is part of the optimization of Maersk Algeria's container backlog management process, through the implementation of a clustering model allowing the segmentation of customers and containers. The objective of this model is to offer a better management of customers and containers by taking into consideration the various factors influencing the retention of containers and their consequences. This solution will allow the company to be more responsive in managing longstandings containers as well as save time by performing differentiated actions on each cluster formed.

Keywords : Longstandings containers, Decision support, machine learning, dashboards.

Résumé :

Ce travail s'inscrit dans le cadre de l'optimisation du processus de gestion des conteneurs en souffrance de Maersk Algérie, par la mise en place d'un modèle de clustering permettant la segmentation des clients et des conteneurs. L'objectif de ce modèle est d'offrir une meilleure gestion des clients et des conteneurs en prenant en considération les différents facteurs influant la rétention des conteneurs et leurs conséquences. La solution en question va permettre à l'entreprise une meilleure réactivité dans la gestion des conteneurs en souffrance ainsi qu'un gain de temps en effectuant des actions différencié sur chaque cluster formé.

Mots clés : Conteneurs en souffrance, aide à la décision, apprentissage automatique, tableaux de bords.

Table des matières

- Table des figures 12
- Liste des tableaux 13
- Liste des abbréviations 14
- Introduction 14
- 1 Présentation de Maersk et du secteur du transport maritime 17**
 - 1.1 Marché mondial du transport maritime : 18
 - 1.2 Marché national du transport maritime 20
 - 1.3 Présentation de Maersk monde. 21
 - 1.3.1 Activités de Maersk : 21
 - 1.3.2 Historique de l'entreprise : 22
 - 1.3.3 Performance financière de Maersk 23
 - 1.4 Présentation de Maersk Algérie 24
 - 1.4.1 Structure organisationnelle 24
 - 1.4.2 Customer Experience 25
 - 1.5 Synthèse 28
- 2 Les conteneurs en souffrance 29**
 - 2.1 Généralités 30
 - 2.1.1 Conteneurs et conteneurisation 30
 - 2.1.2 Types de conteneurs : 32
 - 2.2 Standard operation procedure (SOP) 34
 - 2.3 Analyse de la problématique QQQQCCP 35
 - 2.4 Enseignements tirés & problématique 42
 - 2.5 Synthèse 43
- 3 IA, Apprentissage et Tableau de bord 44**
 - 3.1 Intelligence artificielle 45
 - 3.1.1 Machine Learning 45
 - 3.2 Types d'apprentissage 47
 - 3.3 Réduction dimensionnelle 48
 - 3.4 Partitionnement de données (clustering) 48
 - 3.5 Les étapes du clustering 49
 - 3.5.1 Préparation des données 49
 - 3.5.2 Choix de l'algorithme 50

3.5.3	Validation des modèles	54
3.5.4	Présentation du langage de programmation python	59
3.5.5	Bibliothèques python au service du machine learning	60
3.6	Performance et tableaux de bords	61
3.6.1	La performance	61
3.6.2	Tableaux de bord	61
3.6.3	Méthodologies de construction d'un tableau de bord	63
3.6.4	Technologies de visualisation :	65
3.7	Synthèse	68
4	Segmentations, clustering et Tableau de bord	69
4.1	La démarche CRISP-DM	71
4.2	Segmentation des conteneurs	71
4.2.1	Compréhension du cas	71
4.2.2	Compréhension des données	72
4.2.3	Préparation des données	76
4.2.4	Modélisation des données	81
4.2.5	Evaluation des modèles	84
4.3	Segmentation des clients	86
4.3.1	Compréhension du cas	86
4.3.2	Compréhension et préparation des données	87
4.3.3	Modélisation des données	89
4.3.4	Evaluation des modèles	89
4.4	Déploiement	91
4.4.1	Classification ABC	91
4.4.2	Résultats du clustering	93
4.4.3	Mode opératoire	96
4.5	Mise en place du tableau de bord :	97
4.5.1	Choix de la démarche de construction du tableau de bord	97
4.5.2	Déroulement de la méthode GIMSI :	97
4.6	Synthèse	102
	Conclusion	103
	Bibliographie	105
	Annexes	107

Table des figures

1	Développement du commerce maritime international et de la production mondiale, 2006-2020 (Variation annuelle en pourcentage)	18
2	Le commerce international par région ,2019.(Part en pourcentage du tonnage total)	19
3	Les parts de marché en pourcentage des principaux exploitants de porte-conteneurs dans le monde ,2016	20
4	L'historique de l'entreprise Maersk monde.	22
5	La répartition des revenus Maersk par activité	23
6	Champs géographique de Maersk Algérie	24
7	La structure organisationnelle de Maersk Algérie.	25
8	Organigramme de l'organisation au sein du département Customer eXperience "CX" de Maersk Algérie.	26
9	L'origine de l'apparition des conteneurs	31
10	Les différents types des conteneurs	32
11	Modélisation BPMN de la SOP	34
12	Répartition des conteneurs retenus au niveau des ports opérationnels en Algérie	37
13	Evolution du nombre de conteneurs en rétention	38
14	Répartition des pertes par port	40
15	Les différentes raisons de la non restitution des conteneurs immobiles selon l'extraction de mars 2021.	41
16	Schéma explicatif du fonctionnement d'un programme machine learning	47
17	Schéma représentant les types d'apprentissage automatique	48
18	Processus de clustering	49
19	Les étapes de préparation de données	50
20	Exemple de clustering avec Means Shift	52
21	Exemple de dendrogramme	53
22	Comparaison des résultats obtenus par DBSCAN et K-Means	54
23	Les étapes d'évaluation d'un modèle de clustering	54
24	Exemple illustrant la méthode du coude	56
25	Exemple illustrant la méthode gap statistics	56
26	Les distances a et b sur un exemple de clustering	58
27	Capture d'écran, Google trends	60
28	La performance cercle vertueux entre résultats et moyens consacrés.	61
29	La pyramide d'Ansoff : les trois types de décisions en entreprise.	62
30	Les étapes de la méthode GIMSI	64

Table des figures

31	Le déroulement de la méthode OVAR.	65
32	Figure illustrant les différents outils informatiques de visualisations.	66
33	Interface Graphique d'un TDB sous Tableau	66
34	Interface Graphique d'un TDB sous Zoho Analytics	67
35	Interface Graphique d'un TDB sous Microsoft Power BI.	67
36	Interface Graphique d'un TDB sous SISENCE.	68
37	Diagramme de GANTT du projet	71
38	Schéma des étapes de la démarche CRISP-DM	71
39	Visualisation des variables et leurs types sur python	72
40	Diagramme en camembert représentant la répartition des types de variables	73
41	Heatmap montrant les valeurs manquantes.	74
42	Boite à moustaches sur la variables 'Days'	75
43	Capture d'écran du code python	75
44	Résultats du test de hopkins effectué sur python	76
45	Modalités de la variable Standrad comments	77
46	Modalités de la variable Last Loc Name	77
47	Méthode de sélection des variables	78
48	Variables sélectionnées	79
49	Exemple du label encoding	80
50	Exemple du one hot encoding	80
51	Résultat obtenus suite à la variation du coefficient de silhouette	81
52	Graphique de l'inertie en fonction du nombre de clusters	81
53	Résultat obtenus suite à la variation du coefficient de silhouette en utilisant la CAH	83
54	Visualisation du dendrogramme	83
55	Visualisation des résultats d'évaluation par le diagramme en bâton.	86
56	Visualisation des informations sur la base de données clients	87
57	Visualisation des valeurs manquantes sur python	88
58	Boite à moustaches sur la variable Count of Equipment Number	88
59	Capture d'écran python des variances pour les variables clients	89
60	Fonctions python utilisés pour le calcul des indices d'évaluation	90
61	Scatter-matrix après avoir effectuer le clustering par K-means	91
62	Capture d'écran python des modalités de la variable Standard Comments .	92
63	Schéma expliquant la provenance des données d'entrées pour la classification ABC	92
64	Capture d'écran python montrant le code implémenté pour la classification ABC	92
65	Diagramme en camembert représentant la répartitions des conteneurs sur les clusters	93
66	Schéma récapitulatif de la solution	96
67	Modélisation BPMN du processus mis en place	97
68	Les critères d'un bon indicateur définis par la démarche GIMSI	99
69	L'interface Power BI du tableau de bord élabore	101
70	Modélisation BPMN du processus d'export	108
71	Modélisation BPMN du processus dimport	109

72	Modélisation BPMN du processus d'import	110
73	Modélisation BPMN du processus d'élaboration du manifest	111
74	Modélisation BPMN du processus d'échange documentaire	112
75	Exemple d'une demande de cotation effectuée sur maersk.com.	113
76	Trame indiquant le déroulement des entretiens.	114
77	Scatter-matrix montrant les résultats du clustering obtenus par K-Means.	115

Liste des tableaux

1	La plaque signalétique de Maersk Algérie.	24
2	Récapitulatif des différentes catégories, types et spécifications de conteneurs	33
3	La démarche QQQQCCP adoptée à notre diagnostic	35
4	Les répartitions de restitutions par commentaire selon la base de données de l'année 2021.	38
5	Tableau contenant le prix du fret maritime pour un temps de transit égal à 31 jours	39
6	Tableau des résultats obtenus suite à l'évaluation par le coefficient de silhouette.	84
7	Tableau des résultats obtenus suite à l'évaluation par l'indice de Calinski-Harabasz.	84
8	Tableau des résultats obtenus suite à l'évaluation par l'indice de Davies-Bouldin.	85
9	Tableau des résultats d'évaluation.	85
10	Tableau contenant le nombre de clusters introduit ou obtenu lors de la phase de modélisation.	89
11	Tableau montrant les résultats d'évaluation du modèle de clustering pour les clients	90
12	Tableau récapitulatif du résultat de clustering des conteneurs.	94
13	Tableau récapitulatif du résultat de clustering des clients	95
14	Tableau montrant l'estimation des pertes pour chaque cluster de conteneurs	96

Liste des abbréviations

ACP	<i>Analyse en Composantes Principales</i>
AP	<i>Arnold Peter</i>
BPMN	<i>Business Process Management Notation</i>
BI	<i>Business Intelligence</i>
CM	<i>Case Management</i>
CNUCED	<i>Conférence des nations unis pour le commerce et le développement</i>
CRISP-DM	<i>Cross Industry Standard Process for Data Mining</i>
CS	<i>Customer Service</i>
CX	<i>Customer experience</i>
DBSCAN	<i>Density-based spatial clustering of applications with noise</i>
DUC	<i>Dansk Undergrunds Consortium</i>
DWH	<i>Data Warehouse</i>
ETL	<i>Extraction Transformation Load</i>
GZAL	<i>Grande Zone Arabe de Libre-échange</i>
HR	<i>Human Ressources</i>
IA	<i>Intelligence artificielle</i>
IT	<i>Information Technologies</i>
KPI	<i>Key Performance Indicator</i>

PDG	<i>Président Directeur Général</i>
PL	<i>Party Logistics</i>
SOP	<i>Standard Operating Procedure</i>
TDB	<i>Tableau de bord</i>
TEU	<i>Twenty-foot Equivalent Unit</i>

Introduction

Reflétant l'évolution de l'économie mondiale et de l'activité commerciale, le transport maritime est le mode de transport le plus utilisé pour le commerce international. D'après la CNUCED (Conférence des Nations Unies pour le Commerce Et le Développement), 80% du commerce mondial en volume se font par voie maritime.

Grâce à son expansion très marquée, sa flexibilité et son coût très attractif, le transport maritime a toujours été, et reste encore, le moyen de transport le plus économique. Il constitue donc un instrument privilégié des échanges internationaux et a connu plusieurs révolutions pour s'adapter au fil du temps à l'évolution des échanges.

Les échanges maritimes, portés par les échanges commerciaux internationaux, sont en plein essor. Les flottes de bateaux et de navires appartenant à des centaines de compagnies assurent quotidiennement le transport d'énormes quantités de produits entre différents pays, le transport maritime est donc le moyen le plus efficace pour transporter de grands volumes de marchandises.

Plusieurs facteurs, comme le développement des conteneurs et l'augmentation de la capacité des navires qui sont devenus gigantesques ou encore la multimodalité ont rendu possible cette hausse des échanges maritimes .

Aujourd'hui, la quasi-totalité du commerce international en Algérie est réalisé par voie maritime. Le transport maritime domine les échanges commerciaux avec l'extérieur, il est considéré comme un levier important dans sa politique économique car il a une contribution vitale dans le développement du pays. Cependant le blocage, la non disponibilité et la pénurie des conteneurs constituent des enjeux cruciaux auxquels est confrontée l'industrie maritime notamment avec les conséquences de la crise sanitaire du nouveau coronavirus qui a compliqué beaucoup plus l'acheminement des marchandises.

Les conteneurs en souffrance peuvent être un effet de levier de croissance important pour les industries maritimes. Elles doivent absolument améliorer et maîtriser le pilotage de ses conteneurs pour réagir face aux aléas et assurer un niveau de flexibilité face aux différentes situations.

Selon Maersk, *"Un conteneur est en souffrance dès lors qu'il est immobilisé au niveau du port ou non restitué par le marchand après une période de 22 jours à compter de la date d'accostage du navire"*.

Dans ce contexte, la compagnie maritime sur laquelle a porté notre étude, vise à consolider sa croissance en récupérant des conteneurs dans l'intention d'augmenter leurs rotations effectives pour les rentabiliser.

Afin de répondre aux besoins de la compagnie, ce travail s'intéresse particulièrement à la mise en place de nouveaux outils permettant une gestion plus optimale des conteneurs en souffrance.

Ce qui nous amène à nous poser cette question centrale :

Quels outils pourrait-on mettre en place pour optimiser le processus de gestion des conteneurs en souffrance ?

De ce fait, nous avons formulé les deux hypothèses suivantes :

1. Maersk ne posséderait pas tous les outils nécessaires pour piloter sa performance dans la gestion des conteneurs en souffrance.
2. Les techniques de machine learning ainsi que les tableaux de bords pourraient contribuer à l'amélioration des outils déjà présents au service de la gestion des conteneurs en souffrance.

Pour tenter de répondre à notre problématique nous avons adopté la démarche suivante :

- En premier lieu, nous avons fait un état des lieux à travers des entretiens effectués au sein de l'entreprise, en particulier avec les membres de l'équipe customer experience.
- Compte tenu de la pauvreté des références bibliographiques traitant le problème des conteneurs en souffrance, nous avons consulté des experts de l'entreprise en les soumettant à des vagues successives de questionnements pour générer des avis qui permettent de consolider nos orientations.
- Après avoir extrait les bases de données nécessaires aux préconisations que nous avons élaborées, nous avons eu recours à des logiciels spécifiques tel que Python, Power BI pour analyser les données et mettre en place les modèles mathématiques nécessaires à la segmentation.

Pour but de présenter le travail réalisé par nos soins, décrire les étapes par lesquelles nous sommes passés et présenter les résultats obtenus, nous avons structuré ce document en quatre parties organisées ainsi :

Le premier chapitre sera dédié à la découverte de la compagnie maritime Maersk, présentant d'abord le secteur du transport maritime, son positionnement et ses activités. Déclinant jusqu'à Maersk Algérie, pour éclairer la structure de son entité et de son département "CUSTOMER EXPERIENCE CX " dans lequel nous avons mené notre étude.

Le second chapitre sera consacré à la description du contexte de ce projet et la présentation de la problématique. Il ambitionne de mieux cerner la problématique en s'appuyant sur une étude de l'existant.

Le troisième chapitre vise à vulgariser les différents concepts liés au traitement de données, algorithmes de Machine Learning pour le clustering, leur évaluation ainsi que les méthodes de conception des tableaux de bords.

Le quatrième et dernière chapitre comporte, pour sa part, le déroulement des étapes de la démarche visant à la conception modèle permettant la segmentation des clients et des conteneurs, de par l'utilisation des techniques du Machine Learning et de l'intelligence artificielle, puis de son implémentation dans l'entreprise.

Chapitre 1

Présentation de Maersk et du secteur du transport maritime

Depuis plusieurs années, le transport maritime s'est imposé comme le premier mode de transport de marchandises sur la scène internationale.

Les compagnies maritimes innovent au quotidien pour mieux comprendre leurs clients, anticiper leurs besoins, optimiser l'acheminement de leurs marchandises tout en veillant à limiter leurs impacts environnementaux. Maersk est l'une de ces compagnies maritimes leaders de parts de marché. Elle est solidement installée à la place n°1 sur le podium mondial grâce aux services qu'elle offre à ses clients permettant un développement serein de leurs activités.

Dans ce chapitre, nous allons présenter en premier temps le marché mondial maritime, ainsi que les principaux pionniers du transport maritime national et international. Ensuite, nous présenterons un de ces derniers, le groupe danois AP Møller Maersk premier armateur mondial pour le transport des conteneurs.

Nous exposerons en outre les différentes activités que propose le groupe, ainsi que ses chiffres clés de performance financières. Enfin, nous présenterons la structure organisationnelle de Maersk Algérie et nous nous focaliserons sur le département "Customer eXperience CX" dans lequel nous avons effectué notre stage.

1.1 Marché mondial du transport maritime :

Avec près de 90% du trafic mondial de marchandises, le transport maritime domine les échanges du commerce international. Aujourd'hui 50.000 navires sillonnent les mers : des porte-conteneurs, des vraquiers ou des pétroliers, naviguent à travers la planète, parmi lesquels près de 40% de pétroliers.

Selon la revue du transport maritime 2020, le développement du commerce maritime est représenté par la figure ci-dessous :

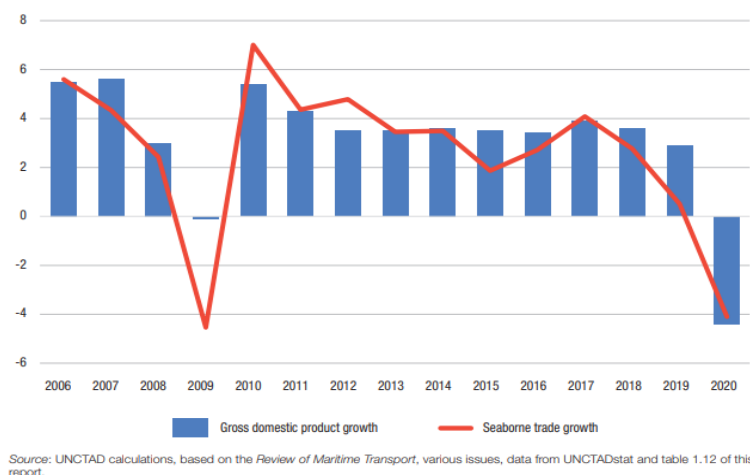
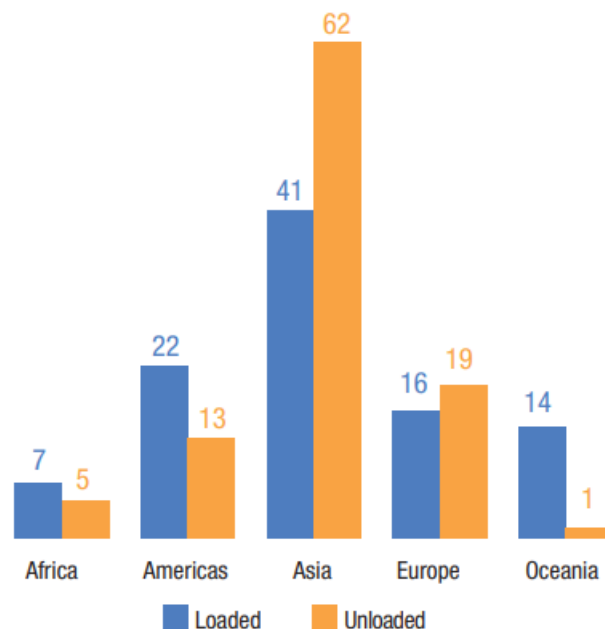


FIG. 1 : Développement du commerce maritime international et de la production mondiale, 2006-2020 (Variation annuelle en pourcentage)

En 2008, nous remarquons de fortes baisses de la production maritime conséquence directe de la crise financière mondiale de 2007. Cette chute spectaculaire se traduit par le déclin du commerce mondial fin 2008-début 2009 de plus de 10%, qui est due principalement selon les raisons évoquées par les chercheurs d'une baisse soudaine de la demande et de l'offre. L'activité maritime s'est fortement détériorée au printemps et au début de l'été 2020 en raison des restrictions aux déplacements mises en place à cause de la pandémie COVID-19. Plusieurs facteurs ont pesé sur la performance du commerce maritime, parmi ces derniers :

1. La pandémie du coronavirus et son impact sur le monde entier.
2. Les tensions en matière de politique commerciale.
3. Conditions économiques défavorables dans certains pays.
4. Les sanctions et la perturbation de l'offre.



Source: UNCTAD calculations, based on data supplied by reporting countries and as published on government and port industry websites, and by specialist sources.

FIG. 2 : Le commerce international par région ,2019.(Part en pourcentage du tonnage total)

Au regard des grands échanges du transport maritime conteneurisé, l'Asie est au cœur du système . En 2019, 41% du tonnage total des marchandises chargées (exportées) provenaient de l'Asie , ainsi que 62% des marchandises déchargées (importées) ont été reçue par cette même région . En effet, sa façade maritime très active dans la mondialisation des échanges maritimes en fait d'elle un continent central et dynamique connecté à l'économie mondiale.

D'un point de vue commercial, le transport maritime de est dominé par seulement quelques groupes internationaux, représentés dans la figure ci-dessous avec leur parts de

marché de l'année 2016.

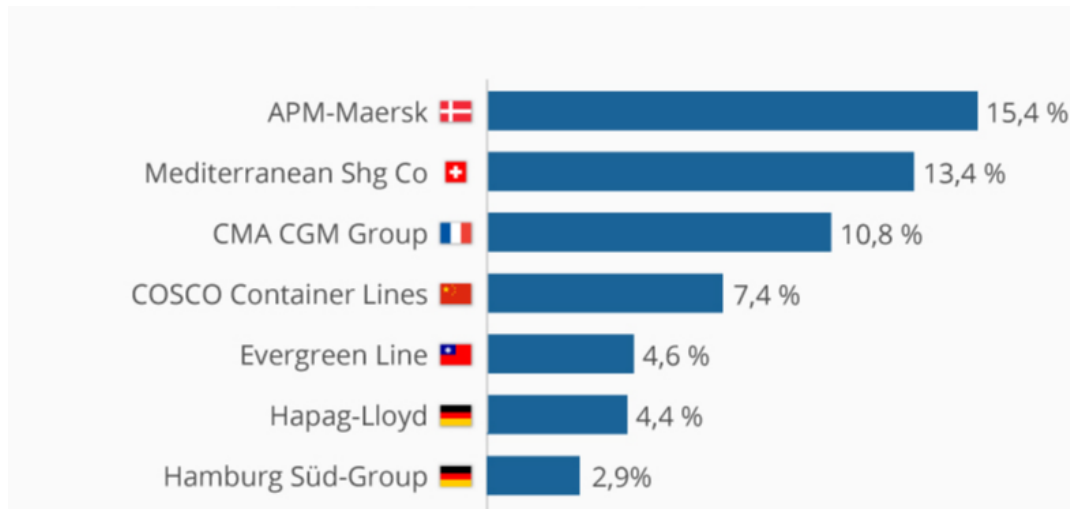


FIG. 3 : Les parts de marché en pourcentage des principaux exploitants de porte-conteneurs dans le monde ,2016

Nous remarquons que le marché mondial maritime est dominé par le groupe APM.Maersk avec plus de 15% du total de parts de marché mondiale. Suivi par MSC, Mediterranean Shipping Company ayant 13,4% parts de marché et CMACGM group avec 10,8%. Ces trois exploitants de porte-conteneurs se distinguent des autres et totalisent ensemble près de 40% des parts de marché, mais aussi ils disposent des flottes les plus importantes au monde.

1.2 Marché national du transport maritime

L'Algérie possède une façade maritime de plus de 1600KM, sa position en fait un acteur stratégique favorisant ses échanges internationaux. Prés de 95% de son commerce international emprunte la voie maritime, son trafic se caractérise par un déséquilibre total entre les importations et les exportations .En termes de trafic de conteneurs, l'Algérie reste un importateur majeur, avec plus de 850 000 Teus, contre 50 000 Teus d'exportations en provenance des ports algériens.

Elle a fait un effort exceptionnel durant la dernière décennie en matière de réalisation de grands projets d'infrastructures et d'équipements publics afin de rendre ce secteur plus performant et plus efficace dans sa contribution dans le développement économique du pays.

De grandes compagnies maritimes européennes sont présentes sur ces routes ,l'offre de services est dominée par la compagnie maritime française "CMA-CGM", Suivi du suisse MSC et en troisième position de Maersk Algérie qui détient à elle seule 20% de parts de marché.

Ces derniers assurent une meilleure accessibilité aux marchés mondiaux, notamment asiatiques, relie l'Algérie à ses partenaires commerciaux méditerranéens tout en tenant en compte des évolutions en cours visant à améliorer la planification de l'infrastructure

portuaire.

1.3 Présentation de Maersk monde.

Maersk est une multinationale danoise créée en 1904 par AP.Moller, avec une équipe dévouée de plus de 100 000 employés, opérant dans environ 130 pays à travers le monde. Il s'agit de la plus grande entreprise au niveau du Danemark et le leader mondial dans le transport par conteneurs avec un chiffre d'affaires s'élevant à 40,311 Md(en2015), 39,01Md (en 2018).

Pendant la plus grande partie du 20ème siècle Maersk était un conglomérat présent dans divers secteurs d'activités (Services informatiques, Banques, Commerce du détail et exploitation pétrolière), l'ensemble de ses activités étaient gérés de manière indépendante.

Au cours des années 2000 Maersk a cédé une grande partie de ses entreprises surtout en 2016 avec la nomination du nouveau PDG Soren Skou ou ils ont décidé de se concentrer sur deux domaines principaux : Transport maritime et Énergies.

1.3.1 Activités de Maersk :

Maersk offre plusieurs services dans le domaine du transport et de la logistique afin de répondre aux besoins sans cesse changeants de ses clients.

Ils s'investissent dans de nouveaux services et de nouvelles technologies pour les secteurs où les défis et les perturbations sont les plus importants.

Grâce à une offre simple de produits, de services numériques, à un engagement transparent et un réseau de transport de qualité, Maersk permet à ses clients de commercer et de se développer en transportant des marchandises partout dans le monde entier. **Maersk est présente sur 4 segments d'activités :**

1. **Ocean** : Au cœur de la stratégie de Maersk, Ocean offre aux clients la flexibilité et la stabilité nécessaires pour gérer et simplifier les chaînes d'approvisionnement de bout en bout. Exploitant la plus grande flotte de porte-conteneurs au monde, elle permet de transporter plus de 12 millions de conteneurs équivalent à quarante pieds par an. Elle s'étend à livrer à travers les quatre coins du globe faisant escale dans plus de 300 ports du monde entier.
2. **Logistics Services** : Les services logistiques constituent l'élément de croissance essentiel de la stratégie Maersk. Ils visent à répondre davantage aux besoins des clients à chaque étape de leur chaîne d'approvisionnement grâce à des offres de logistique intégrée.

Elles sont activées par des plate-formes numériques, comprenant le transport terrestre par camion et par rail, la consolidation/consolidation et la réalisation d'entrepôts avec des services de distribution, des services de courtage

en douane, des services d'expédition aérienne, des services de chargement de conteneurs, l'assurance des marchandises ce qui permet de proposer des services de quatrième partie logistique "services 4PL".

- 3. Terminals Towage** : Ils comprennent les activités d'exploitation de terminaux dans les ports entièrement ou partiellement contrôlés par la marque APM Terminals, dont la principale source de revenus étant les activités portuaires, et les activités de remorquage sous la marque Svitzer, fournisseur de services de remorquage et de sauvetage en mer.
- 4. Manufacturing Others** : Il s'agit de l'ensemble des activités de fabrication assurées par des fournisseurs distincts : Maersk Container Industry qui est le fabricant des conteneurs frigorifiques, Maersk Supply Service qui fournit des services maritimes et des solutions intégrées au secteur de l'énergie dans le monde entier et Maersk Training un fournisseur de services de formation pour le secteur maritime, pétrolier, de l'énergie et gazier.

1.3.2 Historique de l'entreprise :

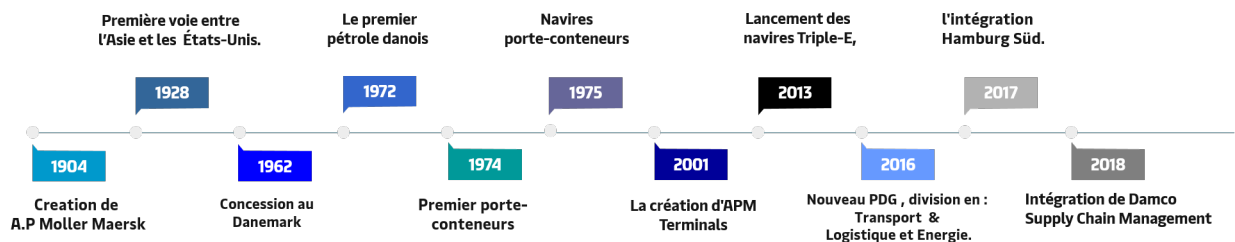


FIG. 4 : L'histoire de l'entreprise Maersk monde.
(Document interne, Maersk)

- **1904** : La compagnie a été créée en 1904 par le capitaine Peter Mærsk Møller et son fils Arnold Peter.
- **1928** : Le premier service régulier de ligne avec des départs mensuels a été lancé en 1928 entre l'Asie et les États-Unis.
- **1962** : A.P. Møller et les compagnies de navigation ont obtenu une concession exclusive pour l'exploration et l'extraction de matières premières du sous-sol danois pour une période de 50 ans.
- **1972** : La coentreprise Dansk Undergrunds Consortium (DUC) a fait sa première découverte et a commencé à produire le pétrole dans le champ Dan à l'été 1972 .
- **1974** : Maersk a fait l'acquisition de son premier porte-conteneurs. L'armement devient par la suite l'un des plus gros armements de porte-conteneurs du monde.
- **2001** : Maersk a investi dans des installations de terminaux dans le monde entier.

- **2013** : le premier porte-conteneurs Triple-E a été livré. Les caractéristiques de conception uniques de la série Triple-E garantissent des navires plus efficaces sur le plan énergétique, des économies d'échelle, tout en réduisant considérablement les émissions de CO2 par conteneur.
- **2016** : Søren Skou a été nommé directeur général d'A.P. Moller - Maersk en 2016. Il travaille dans l'entreprise depuis 1983 et est membre de l'équipe de direction d'A.P. Moller - Maersk depuis 2001.
- **2017** : Fin 2016, A.P Moller a annoncé son intention d'acquérir la compagnie allemande de transport maritime par conteneurs Hamburg Süd.
- **2018** : Maersk Line et Damco Supply Chain Management annoncent leur intégration sous le nom de Maersk.

1.3.3 Performance financière de Maersk

Les revenus respectifs des activités citées auparavant sont présentés par le graphique suivant :

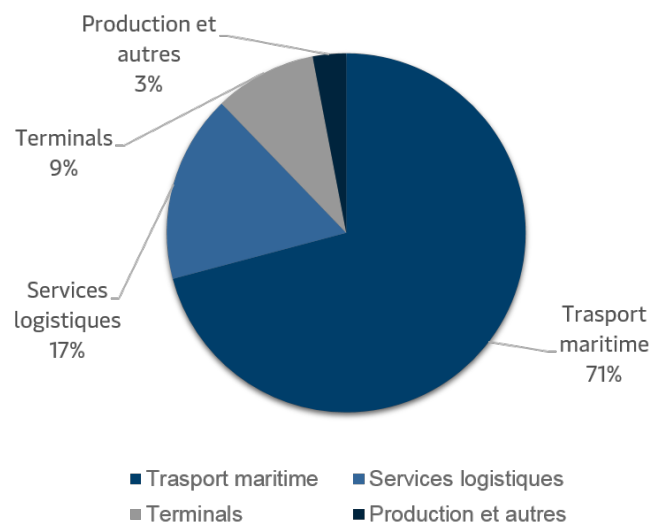


FIG. 5 : La répartition des revenus Maersk par activité

Nous avons pu constater que plus de 70% des revenus cumulés durant l'année 2020 ont été collectés uniquement grâce à l'activité "OCEAN" de transport maritime. En effet, cette dernière représente une activité clé pour l'entreprise permettant d'améliorer ses bénéfices et son flux de trésorerie. Les trois autres activités accumulent environ 30% et complètent considérablement les revenus de l'entreprise malgré toutes les conditions liées à la crise sanitaire .

1.4 Présentation de Maersk Algérie

Maersk s'est installée en Algérie en 2000, avec des agences à proximité des cinq principaux ports algériens et un siège social à Alger, Maersk s'est bien placée pour aider à propulser les entreprises vers une plus grande croissance. Elle assure des liaisons régulières au départ des cinq principaux ports opérationnels : Alger, Annaba, Béjaia, Oran et Skikda pour connecter les entreprises à leurs marchés clés partout dans le monde.

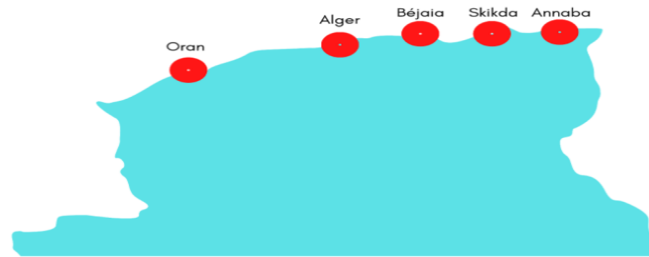


FIG. 6 : Champs géographique de Maersk Algérie

Avec son expertise, ses services adaptés aux besoins de tous les types des cargaisons, Maersk est l'un des acteurs principaux sur le marché Algérien. Quels que soient les secteurs d'activités, les marchandises, elle propose des solutions contribuant au développement de toutes les entreprises - petites ou grandes mais aussi permettant de relier les entreprises au monde entier et simplifier leurs expéditions.

Les données respectives de Maersk Algérie sont résumées dans le tableau ci-dessous :

Nom de l'entreprise	Maersk Algérie
Date de création	29/07/2000
Statut juridique	SPA
Siège social	Hydra, Algérie
Le capital social	30,000,000.00 DZD
Le chiffre d'affaires (CA)	1,796,370,831 DZD
Le nombre de salariés	43

TAB. 1 : La plaque signalétique de Maersk Algérie.

1.4.1 Structure organisationnelle

Le schéma organisationnel ci-dessous est fait en sorte de prendre en charge les différentes fonctions de Maersk Algérie, ainsi que leurs missions sont résumées ci-après :

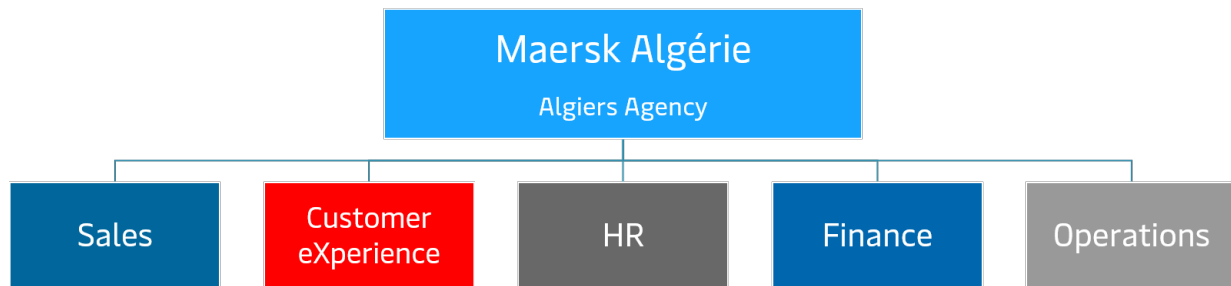


FIG. 7 : La structure organisationnelle de Maersk Algérie.

- **Département Sales** : il s'agit du commercial, sa mission principale consiste à élargir son portefeuille clients grâce à des techniques de vente et de prospection et ainsi développer les ventes de son entreprise et remplir son objectif de chiffre d'affaires.
- **Département Finances** : Ce département se situe au cœur de l'entreprise, il se charge du suivi des coûts, la supervision de la comptabilité et la gestion de la trésorerie mais aussi la vérification de la rentabilité, solvabilité de l'entreprise. De manière très large, cette fonction doit connaître son marché : l'évolution des concurrents, des fournisseurs et de tous les acteurs qui gravitent autour de l'entreprise.
- **Département IT** : Il a la responsabilité de garder la technologie informatique de l'organisation en marche. Ainsi l'équipe IT est responsable du matériel, des logiciels et de la mise en réseau des ordinateurs de l'entreprise.
- **Département HR** : Ce département s'occupe de la gestion des ressources humaines, des recrutements, des formations, l'administration de la paie ainsi que les conditions de travail et les événements.
- **Département Customer Experience "CX"** : Le département dans lequel nous avons réalisé notre projet, afin de comprendre son rôle et son fonctionnement nous allons le présenter par la suite.

1.4.2 Customer Experience

Le département Customer Experience représente l'ensemble des interactions entre le client et l'entreprise tout au long de son parcours. Sa mission s'étend de la réservation

effectuée par le client et la gestion des opérations tout au long de l'expédition jusqu'à la facturation finale et le traitement des retours et réclamations. Même s'il est difficile de gérer et de tenter de contrôler l'expérience d'un client au cours de son parcours, il est très important pour une entreprise d'essayer de le faire.

Le département aura donc la responsabilité d'être proactif à toutes les étapes du parcours client, afin de satisfaire ses attentes et ses besoins. Il s'occupe de réaliser à un rythme régulier des enquêtes de satisfaction clients permettant de mesurer la performance du service.

L'organigramme ci-dessous représente l'organisation du département Customer eXperience :

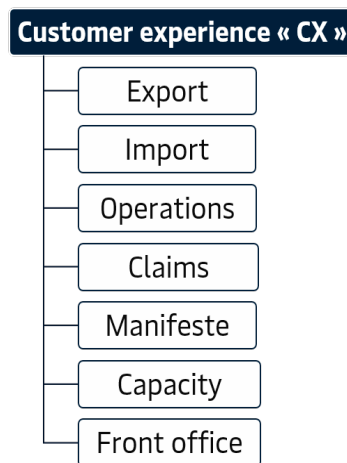


FIG. 8 : Organigramme de l'organisation au sein du département Customer eXperience "CX" de Maersk Algérie.

Services du département :

- **Export** : Ce service s'occupe de la gestion et le suivi des exportations de marchandises nécessaires à l'activité de Maersk Algérie. Il est tenu à informer, coordonner et assister les clients de la réservation sur la plateforme Maersk.com et les demandes de cotations jusqu'à l'arrivée de leurs marchandises au port destiné à l'étranger.
La modélisation du processus d'export est présentée dans l'**Annexe A** p108.
- **Import** : La principale mission de ce service est l'accompagnement des clients au cours de leurs processus d'importation, ainsi que la supervision des tâches qui lui sont liées, à travers :
 - La vérification des informations contenues dans les connaissements.
 - S'assurer que les certificats de non manipulation sont prêts, ces derniers sont des documents élaborés par la douane du port de transbordement suivant la convention des pays arabes "GZALE". Ils permettent au pays conventionnés de bénéficier d'une facilitation d'échanges commerciaux et de réduction des taxes douanières.

- Le suivi de l’envoi des avis d’arrivée, relâchement des bon à délivrer par le front office.

La modélisation du processus d’import est présentée au niveau de l’**Annexe B**.
p109

- **Operations** : Ce service se charge principalement de l’organisation et la gestion des opérations sur le terrain, avec les “Third Party” les sous-traitants de la compagnie maritime. Il est a noter que la planification du baie plan se fait en coordination avec le capitaine du navire en prenant compte des conteneurs de tailles standards spécifiques. Ainsi les plans effectués viseront à maximiser l’économie d’expédition et de sécurité à bord.
- **Claims** : La mission principale de ce service est le traitement des réclamations et retours clients tout en gardant la meilleure image possible de Maersk Algérie. Ces réclamations peuvent concerner : les pertes de cargaisons, les marchandises avariées dues à des cas de force majeure ou des événements indépendants de la volonté des parties responsables.

La modélisation du processus de réclamations est présentée au niveau de **Annexe C** p110.

- **Manifeste** : Service travaillant en coordination avec les douanes Algériennes à des fins réglementaires. Après une vérification rigoureuse, le document “manifeste” doit être envoyé à la douane à l’arrivée du navire dans le port. Ce dernier est une compilation de renseignements sur les marchandises transportées au niveau de chaque navire, Il reprend ainsi, pour toutes les expéditions individuelles, les informations utiles contenues dans les titres de transport :

- L’expéditeur et le destinataire.
- Un numéro de gros lié au manifeste.
- Les numéros, nombres et types de conteneurs.
- La nature d’emballage et la désignation des marchandises.

Une modification du manifeste après sa validation par la douane engendrera des frais supplémentaires à payer.

Le processus d’élaboration du manifeste est présenté au niveau de **Annexe D** p111.

- **Capacity** : Le service s’assure de l’adéquation entre la capacité de chargement disponible dans le navire et les réservations effectuées par les clients. Le chargé “Capacity” tâchera de sélectionner les marchandises à charger au bord du navire selon quelques paramètres de priorisation : types de conteneurs, marchandises et la nature du client et sa relation avec l’entreprise. Dès la saturation du navire, le chargement des conteneurs est arrêté ainsi les conteneurs non chargés seront rollés aux prochains voyages.
- **Front Office** : Ce service représente l’interface de Maersk auprès de ses clients, sa mission du service se répartie selon plusieurs opérations :

- La réception et vérification des documentations présentés par le client physique et sur système.
- Le paiement des charges fixes (d'agence, manutention,...etc) et les avances sur surestaries.¹
- Le traitement de l'ensemble des factures relatives aux prestations délivrées aux clients.
- La remise du bon à délivrer au client.

Le processus d'échange documentaire est présenté au niveau de **Annexe E**. p112

1.5 Synthèse

Dans ce chapitre, nous nous sommes familiarisés avec l'environnement de l'entreprise, son secteur d'activité et sa performance financière ainsi que ses activités et services. Cependant, nous nous sommes focalisé sur la présentation du département Customer experience "CX" où notre étude a été menée, ainsi que son organisation et ses différentes missions.

Dans le chapitre qui suit, nous aborderons le processus de gestion des conteneurs en souffrance pour cerner le contexte de notre étude. Par la suite, nous effectuerons un diagnostic a fin de faire ressortir les dysfonctionnements relatifs à ce dernier et les pistes d'amélioration que nous pourrons exploiter.

¹Lorsque le chargeur n'arrive pas à achever son déchargement ou son chargement dans le délai fixé, il paye un coût supplémentaire : la surestarie (demurrage).

Chapitre 2

Les conteneurs en souffrance

Les conteneurs sont aujourd'hui le principal type d'équipement utilisé dans le transport intermodal. Un conteneur possède une unité de chargement standardisée qui convient aux navires, aux camions et aux trains et qui peut être rapidement transférée d'un mode de transport à l'autre.

Après la crise sanitaire, la chine est devenue le premier partenaire commerciale pour l'Union Européenne, le trafic maritime entre les deux représente 36% du trafic mondiale mais le pays trouve des difficultés pour transporter les cargaisons dans les délais à cause d'une pénurie de conteneurs.

Aujourd'hui la disponibilité des conteneurs représente un défi pour l'ensemble des compagnies maritimes et afin de récupérer leurs conteneurs, Maersk a mis en place une SOP (Standard Operation Procedure) dans le but de récupérer les conteneurs en souffrance et les évacuer.

Dans ce chapitre, nous allons présenter le processus de gestion des conteneurs en souffrance en passant par une description générale du processus accompagnée par une modélisation détaillé de ce dernier, par la suite nous présenterons le diagnostic que nous avons effectué par la méthode QQQQCCP¹ qui nous a permis de comprendre le processus en profondeur.

A la fin de ce chapitre nous présenterons les résultats de ce diagnostic ainsi que la problématique à laquelle nous répondrons à la fin de ce travail.

2.1 Généralités

2.1.1 Conteneurs et conteneurisation

"La conteneurisation constitue aujourd'hui l'épine dorsale de la mondialisation. Elle permet de mettre les marchandises les plus diverses dans des conteneurs, boîtes métalliques aux dimensions standardisées, pendant toute la durée de leur transport maritime et terrestre, de la porte de l'industriel jusqu'à celle du client. Avec la conteneurisation, le transport maritime est intégré dans des chaînes de transport multimodales qui se déploient à l'échelle de la planète au service du commerce international."[Antoine Frémont 2009]

Nous présentons à travers la figure suivante l'origine des conteneurs et de la conteneurisation et son évolution à travers le temps :

¹Série de questionnement permettant l'analyse d'une situation donnée.

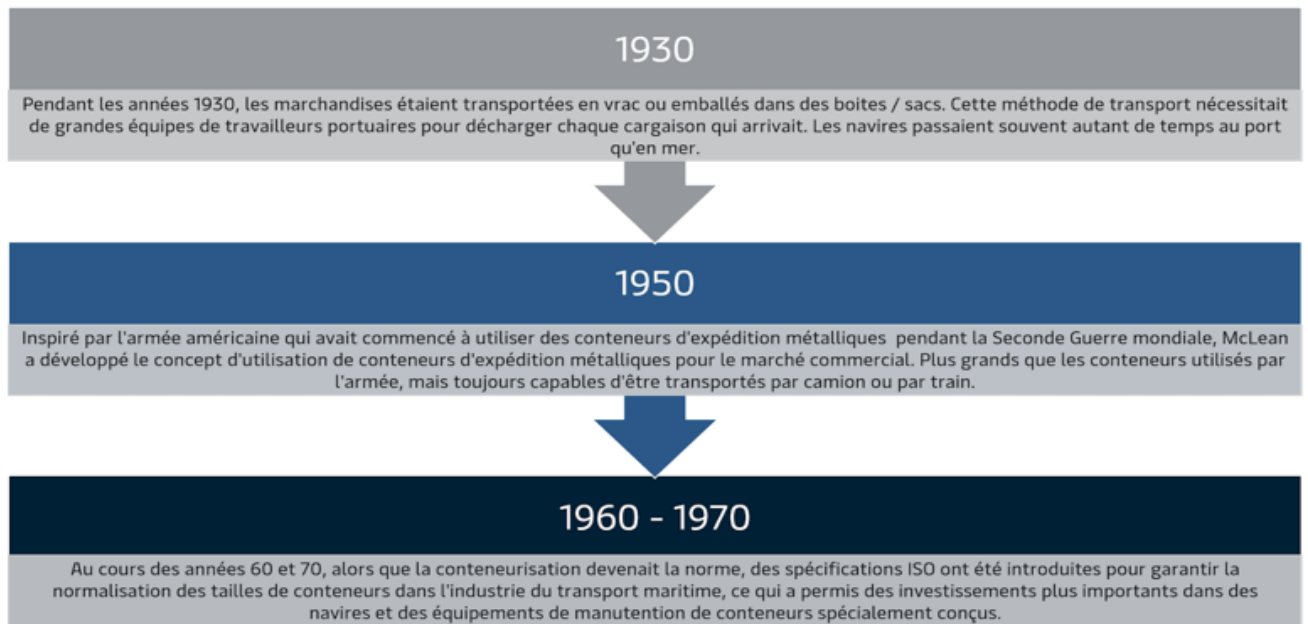


FIG. 9 : L'origine de l'apparition des conteneurs

Aujourd'hui, la conteneurisation a beaucoup contribué au développement du commerce internationale et présente plusieurs avantages :

- **Standardisation** : Le conteneur est un produit de transport standard qui peut être manipulé partout dans le monde (norme ISO) à travers plusieurs modes (navires, camions, barges et wagons), des équipements et des terminaux spécialisés.
- **Flexibilité** : Les conteneurs peuvent être utilisés pour transporter une grande variété de marchandises telles que des produits de base (charbon, blé), des produits manufacturés, des voitures et des marchandises réfrigérées (périssables). Il existe des conteneurs adaptés pour les marchandises sèches, les liquides (pétrole et produits chimiques). Les conteneurs mis au rebut peuvent être recyclés et réutilisés à d'autres fins.
- **Coûts** : Le transport par conteneur permet de réduire les coûts de transport grâce aux avantages de la normalisation. Le transport d'une même quantité de marchandises en vrac dans un conteneur est environ 20 fois moins coûteux que les moyens conventionnels. Les conteneurs permettent de réaliser des économies d'échelle dans les modes et les terminaux, ce qui n'était pas possible avec la manutention standard de marchandises diverses.
- **Vitesse** : Les opérations de transbordement sont minimales et rapides, et les délais de rotation des ports maritimes ont été réduits de 3 semaines à environ 24 heures. Grâce à cet avantage de transbordement, les chaînes de transport impliquant des conteneurs sont plus rapides. Les réseaux de transport par conteneurs sont bien connectés et offrent un large éventail d'options d'expédition.

- **Entreposage** : Le conteneur est son propre entrepôt, protégeant la cargaison qu'il contient. Cela implique un conditionnement plus simple et moins coûteux des cargaisons conteneurisées.
- **Sécurité** : Le contenu du conteneur est inconnu des transporteurs puisqu'il ne peut être ouvert qu'à l'origine (vendeur/expéditeur), à la douane et à la destination (acheteur). Cela implique une réduction des pertes et des détériorations (vols).

2.1.2 Types de conteneurs :

Il existe deux types courants de conteneurs, les conteneurs de vingt et de quarante pieds de long, une unité de vingt pieds ou 1 TEU est mesurée extérieurement à environ 6 mètres de long. La capacité d'un navire varie en fonction du nombre de conteneurs de vingt pieds ou de quarante pieds (ou les deux à la fois) présents à bord. Les types de conteneurs d'expédition peuvent varier en fonction du type de marchandises expédiées ou des services spéciaux pour lesquels ils sont nécessaires.

Les unités de conteneurs peuvent varier en termes de dimensions, de structure, de matériaux et de construction. Certains des conteneurs les plus couramment utilisés sont mentionnés ci-dessous :



FIG. 10 : Les différents types des conteneurs

1. **DRY CARGO** : Le type de conteneur le plus courant au monde. Il est adapté au transport de marchandises sec tel que les produits de grande distribution.
2. **Reefer Cargo** : Ce type de conteneurs est capable de maintenir une température fixe quelles que soient les conditions climatiques externes, ils servent à transporter les produits frais, surgelés et pharmaceutiques.
 - Des conteneurs frigorifiques ou des conteneurs réfrigérés sont utilisés pour transporter des produits tels que les fruits, légumes, produits laitiers et viande sont des marchandises typiques à expédier dans ce type de conteneur maritime.

3. **SPECIAL CARGO** : Les marchandises surdimensionnées et lourdes dépassent les proportions et les capacités d'un seul conteneur standard, Maersk propose des solutions pour transporter les marchandises spéciales :

- **Marchandise in Gauge** :

Les dimensions des marchandises sont inférieures ou égales à celles d'un conteneur plate-forme ou à toit ouvrant (elles nécessitent un équipement spécial, mais n'impliquent pas d'emplacements supplémentaires sur le navire).

- **Marchandise hors Gabaris** :

Les dimensions des marchandises dépassent celles du conteneur plate-forme ou à toit ouvrant en longueur, largeur ou poids, mais elles peuvent toutefois être unitarisées dans des conteneurs.

- **Marchandise Diverse** :

Elles dépassent largement la taille et/ou le poids d'un conteneur standard, sont traitées comme des marchandises non unitarisées et sont généralement placées sur une plateforme avec de nombreux points d'arrimage. Exemples de chargements break bulk : yachts, machines industrielles ou wagons de train.

-Les conteneurs Flat rack sont particulièrement adaptés aux charges lourdes et aux cargaisons nécessitant un chargement par le haut ou les côtés, comme les tuyaux et machines.

-Les conteneurs à toit ouvert n'ont pas de toit mais sont recouverts d'une bâche au lieu d'un toit solide. Cela permet de charger, par le haut, des marchandises surdimensionnées telles que du bois et de la ferraille.

	Types conteneurs	Spécifications		
Catégories	size	Max.gross	Tare	Max.payload
Dry Cargo	20' standard	30.480	2.280	28.200
	40' standard	32.500	3.700	28.800
	40' High Cube	32.500	3.880	28.620
	45' High	32.500	3.900	28.600
Refrigerated Cargo	20' standard Reefer	30.480	2.990	27.490
	40' High Cube Reefer	34.000	4.400	29.600
	40' High Cube Magnum Reefer	34.000	4.275	29.725
	40' High Cube StarCare	34.000	4.870	29.130
	40' High Cube-Super Freezer	34.000	4.500	29.500
Special Cargo	20' Open Top	34.000	4.890	29.500
	40' Open Top High Cube	34.000	2.990	29.500
	40' High Flat Rack	34.000	4.500	29.500

TAB. 2 : Récapitulatif des différentes catégories, types et spécifications de conteneurs
(Document interne, Maersk)

2.2 Standard operation procedure (SOP)

Une standard operation procedure est un document facile à comprendre expliquant étape par étape, les flux de travail et les opérations de routine. Elle consiste à systématiser tous les processus de votre entreprise et à les regrouper dans un seul dossier destiné aux collaborateurs concernés.

Voici quelques exemples de SOP que nous pouvons trouver au sein d'une entreprise.

- Enregistrement des nouveaux clients.
- Politique d'une entreprise.
- Instructions de travail.

Pour le cas de la gestion des conteneurs en souffrance, nous avons pu modéliser la SOP correspondante par le langage BPMN² de modélisation :

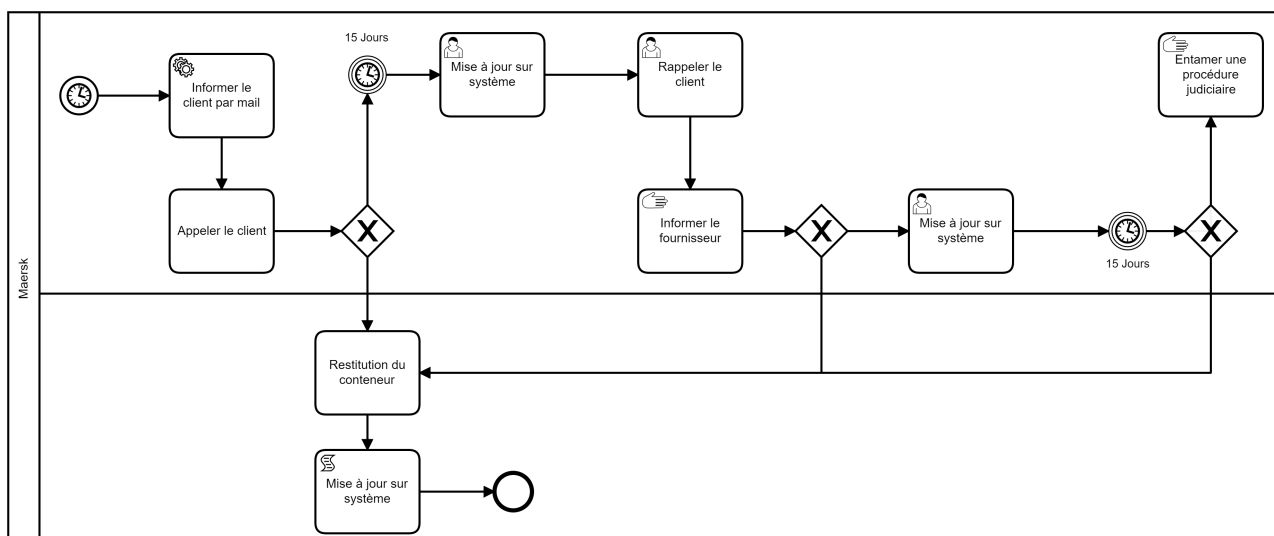


FIG. 11 : Modélisation BPMN de la SOP

La SOP décrite dans la cartographie précédente regroupe les actions qu'il faut entreprendre pour accompagner le client et limiter les conséquences relatives aux conteneurs en souffrance.

- **De 22 à 45 jours :** L'équipe CX est amenée à appeler et envoyer des notifications aux clients concernés par mail. Ces notifications ne peuvent être envoyées qu'à partir de Case Management (CM) de Salesforce. Ces dernières peuvent inciter le client à restituer ses conteneurs dans le plus bref délai possible.
- **De 45 à 60 jours :** La même procédure est appliquée (appels et mails), le fournisseur d'origine est informé par l'état d'avancement de son client. Les retours clients sont mis à jour sur le système .

²Méthode de modélisation de processus d'affaires pour décrire les chaînes de valeur et les activités métier d'une organisation

- **De 60 à 90 jours** : Les mêmes actions sont entreprises, le client est informé que s'il ne restituera pas ses conteneurs avant les 90 jours, une lettre de relance sera envoyée à travers un huissier de justice.
- **Au-delà de 90 jours** : Le cas devient un cas légal, la lettre de mise en demeure est envoyée au client pour se présenter à la justice.

2.3 Analyse de la problématique QQQQCCP

Dans cette partie, nous allons se focaliser sur le processus de gestion des conteneurs immobiliers qui fait l'objet de notre étude. Pour ce faire, nous avons commencé par établir un diagnostic, la méthode QQQQCCP peut être considérée comme l'outil analytique le mieux adapté permettant de faire une analyse fine de la situation et de se poser les bonnes questions pour comprendre les dysfonctionnements et y répondre plus efficacement.

Ce diagnostic est basé sur les informations issues des sources suivantes :

-Des réunions d'échanges avec les différents responsables du service "customer experience"(CX) pour mieux cerner les besoins de l'entreprise sur les différents axes.

-Rapports et documents internes de l'entreprise.

Ce diagnostic favorise l'évaluation de la retenue des conteneurs immobiliers sous ses différentes facettes, en ne négligeant aucun aspect. **Il répond à un double objectif :**

-La détermination des différents facteurs susceptibles d'affecter l'immobilisation des conteneurs, et donc l'activité de Maersk.

-L'identification des faiblesses qui pourraient constituer un frein ou un obstacle pour la restitution des conteneurs, mais aussi les opportunités et les pistes d'amélioration auxquelles un plan d'action stratégique devra être établi pour y faire face.

La démarche consiste à récolter et analyser toutes les informations pertinentes disponibles tirées de la base de données qui représente l'état actuel des conteneurs retenus (Mars 2021).

La démarche adoptée est résumée dans le tableau suivante :

Quoi ?	les conteneurs retenus .
Qui ?	par des clients/la douane .
Où ?	les ports opérationnels d'Algérie.
Quand ?	à partir de 22jours
Comment ?	leur restitution se fait grâce à plusieurs actions
Combien ?	Nombre de conteneurs bloqués ainsi que les pertes engendrées
Pourquoi ?	leur blocage est du à de multiples raisons

TAB. 3 : La démarche QQQQCCP adoptée à notre diagnostic

1. Quoi ?

Comme mentionnée plus avant, Il s'agit des conteneurs en souffrance qui se trouvent bloqués non évacués au niveau des différents ports opérationnels de Maersk Algérie.

Ceci peut être dû à plusieurs sources de problèmes que nous allons expliquer ultérieurement.

Le blocage de ces derniers entraîne une surcharge dans les entrepôts de la douane et un manque à gagner pour la compagnie maritime qui n'arrive pas à faire tourner ses conteneurs mais aussi à les rentabiliser.

2. Qui ?

L'évacuation des conteneurs en souffrance dépend étroitement des clients mais aussi de la douane des ports humides.

- **Le client :**

Le clients est d'une considération importante à prendre en compte dans la gestion des conteneurs en souffrance.

Il représente donc un maillon essentiel dans les procédures à effectuer tout au long des échanges documentaires et douaniers. Cependant, un manque d'engagement de ce dernier pourrait rendre le processus encore plus chronophage.

- **La douane portuaire :**

Chaque conteneur en souffrance comporte un dossier administratif et préalablement un dossier juridique si son séjour dépasse les délais légaux de séjour et souvent ce sont les lourdeurs de ces procédures judiciaires qui prolongent le séjour de ces marchandises.

L'administration douanière avertit que les conteneurs entreposés depuis une longue période, dépassant les délais légaux de séjour seront soumis à la vente aux enchères publiques ou à la destruction de leur contenu pour ceux en butte à des effets d'avarie.

C'est un constat malheureux car nous nous retrouvons avec des marchandises, quelques fois périmées, qui occupent des espaces pour rien et qui attendent toujours leur évacuation, voire leur destruction.

3. Où ?

Nous avons commencé par représenter le graphique illustrant la répartition des conteneurs en souffrance dans les différents ports opérationnels de Maersk selon les données relatives à l'état actuel des conteneurs immobiles.

Les résultats sont les suivants :

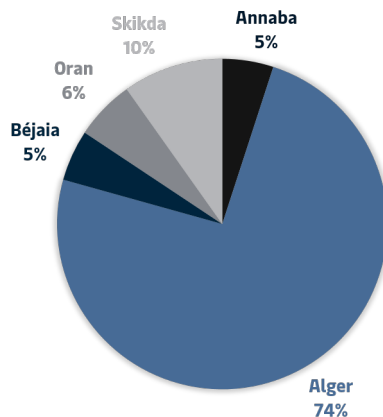


FIG. 12 : Répartition des conteneurs retenus au niveau des ports opérationnels en Algérie

A partir de ce graphique, nous remarquons clairement que :

- L'ensemble des conteneurs retenus se répartit au niveau des cinq ports opérationnels de Maersk Algérie : Alger, Oran, Béjaïa, Annaba et Skikda.
- La majorité écrasante des conteneurs, soit 74% des conteneurs, quasiment les trois quarts sont situés au niveau du port d'Alger. Cette proportion est justifiée par le volume d'activité considérable présent sur Alger en comparant avec les autres ports du pays.

4. Quand ?

On dit qu'un conteneur est en souffrance lorsqu'il n'est pas restitué par le client avant 22 jours qui est la durée moyenne estimée comme durée suffisante pour qu'un client se présente à la douane portuaire et finalise ses procédures d'échange de documentations.

Cependant, pour toute utilisation supplémentaire du conteneur en extra-time, la compagnie maritime exige à son client un certain montant à payer "les frais de surestaries" compte tenu du temps pris par ce dernier pour la restitution du conteneur.

Dans le cas de Maersk Algérie, le nombre de "free days" ³ dépend du type de conteneur et de la négociation du fournisseur de la franchise de son client. Pour un conteneur DRY par exemple, au-delà de quinze jours à compter de l'accostage du navire commence la première tranche des frais de surestaries.

5. Comment ?

Dans ce contexte, à partir du 22ème jour l'entreprise tente de récupérer de son clients en déclenchant la SOP expliquée précédemment.

Cependant, plusieurs actions sont mises en place pour tenter de restituer les

³Les compagnies maritimes accordent un certain nombre de jours au chargeur pour déposer le conteneur. Durant cette période, le chargeur est exonéré des frais de surestarie.

conteneurs immobiles dans les plus brefs délais possibles. Selon la base de donnée relative aux restitutions de l'année 2021, nous avons tiré les données suivantes :

Commentaires	Restitutions 2021
Appels et rappels	412
Accord avec la douane	3
Pression du service réclamation	108
Restitué par le client	1236
Accord avec client	789

TAB. 4 : Les répartitions de restitutions par commentaire selon la base de données de l'année 2021.

La plupart des restitutions effectuées en 2021, voir 1236 conteneurs sont faites par les clients, sans aucune intervention ni pression des équipes Maersk. 789 conteneurs ont été restitués suite à un accord avec les clients qui a permis de libérer les cargaisons, mais seulement 3 conteneurs ont été évacués grâce à un accord avec la douane Algérienne. Les efforts d'appels à plusieurs reprises et l'envoi des e-mails renforcent les restitutions (412 conteneurs), mais aussi le traitement des réclamations des clients soulage la souffrance d'une bonne partie des conteneurs.

6. Combien ?

Le graphique suivant représente l'évolution du nombre de conteneurs en rétention de 2018 jusqu'à 2020 :

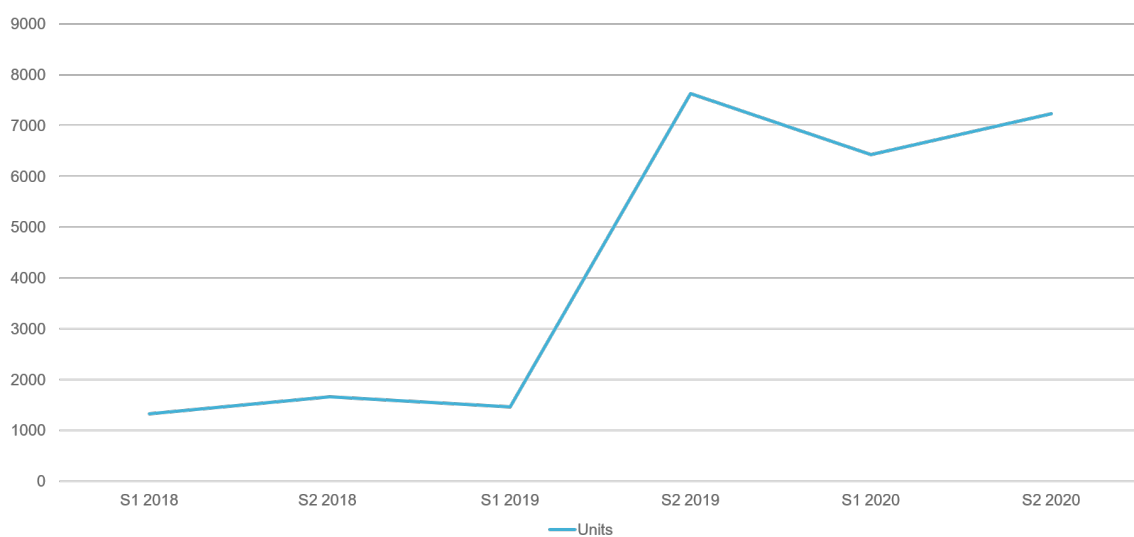


FIG. 13 : Evolution du nombre de conteneurs en rétention

Nous remarquons une augmentation d'environ 419% du nombre de conteneurs, cette augmentation est due aux blocages liés à la situation politique du pays qui a engendré un volume important de marchandises et conteneurs en souffrance.

Dans l'intention d'évaluer les pertes d'opportunités liés à la rétention des conteneurs, nous nous sommes appuyés les informations disponibles au niveau du site officiel de Maersk afin d'effectuer des demandes de cotations (**Annexe F** p113) et obtenir les coûts pour chaque type de conteneurs disponibles en Algérie, pour cela nous avons fixé les paramètres de distance, temps et type de marchandises dans le but d'obtenir des valeurs qui peuvent être comparable.

Nous avons fixé le port de Havre en France comme point de départ et celui d'Alger comme point d'arrivée, avec un type de marchandises " Dry Standard " et pour chaque demande de cotation nous avons varier le type et la taille du conteneurs, le transit time est de 31 jours.

Le tableau suivant représente les prix du fret maritime pour chaque type et taille de conteneurs :

Types	Taille	Coût (EUR)
DRY	20	2610
	40	1856
Reefer	20	2645
	40	2645
Open	20	570
	40	900
Flat	20	570
	40	900

TAB. 5 : Tableau contenant le prix du fret maritime pour un temps de transit égal à 31 jours

Afin d'estimer la perte d'opportunités pour les conteneurs en souffrance que nous avons dans notre base de données, nous supposons que la variation est linéaire par rapport au nombre de jours dans le but de tirer une valeur approximative du coût, la valeur totale de ce dernier est obtenu par la somme des coûts sur l'ensemble des conteneurs présents dans la liste.

Les pertes d'opportunités liés aux conteneurs en souffrance sont estimés à 305 789 489,77 €, le graphique suivant représente la répartition des pertes sur les différents ports ou opère Maersk en Algérie :

■ Algiers Terminal-Algeria ■ Annaba Terminal-Algeria ■ Entreprise Portuaire de Bejaia-Algeria ■ Oran Terminal-Algeria ■ Skikda Terminal-Algeria

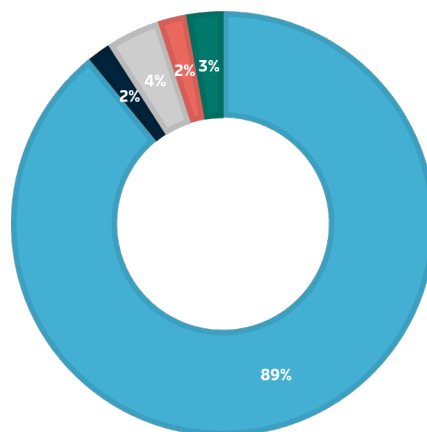


FIG. 14 : Répartition des pertes par port

7. Pourquoi ?

Il est important de recenser toutes les raisons où les conteneurs se retrouvent bloqués afin d'avoir une vision globale sur leur processus de gestion. De ce fait ci-dessous sont présentées l'ensemble des modalités qui sont susceptible de changer considérablement d'un client à un autre, mentionnés commentaires :

- **Cargo abandoned by customer, awaiting for stripping** : Il s'agit d'une cargaison abandonnée par le client, en attente de la destruction ou de la vente aux enchères par la douane.
- **Container cannot be released due to missing OBL⁴ and/or payment from customer** : Ce sont des situations où les clients ont des problèmes de paiement, une documentation non complète (absence du connaissance à titre d'exemple).
- **Container verified in good condition, will be used for export booking** : En faisant les réservations en ligne, les fournisseurs peuvent commettre des erreurs concernant les ports à destination. Les conteneurs se trouvent donc bloqués en attente d'une autre réexportation.
- **Customs documentation delays** : A fin de débloquer leurs marchandises, les clients doivent se rapprocher à temps de la douane portuaire. Tant que l'échange documentaire ne se fait pas dans les délais prévus, les conteneurs restent coincés dans les ports.
- **Incorrectly updated in RKEM (state action plan in free text field)** : L'entreprise n'as toujours pas été notifié par la douane de la destruction/vente aux enchères qui a déjà eu lieux.
- **Legal actions started** : A un stade avancé, une procédure judiciaire est entamée à travers un huissier de justice ainsi le cas devient un cas légal.

⁴Il s'agit du connaissance relatif au transport de marchandises

- **Notification Sent to Customer** : Il s'agit de notifier à chaque fois le client via des mails pour lui expliquer la situation et l'informer qu'il a dépassé les délais prévus.
- **Notification to customer failed** : Les coordonnées client sont incorrectes, par conséquent l'entreprise ne pourra pas contacter le client concerné.
- **Called Customer** : Le client à été déjà contacté par téléphone.
- **Dsp : Lost/ Stolen** : C'est le cas d'un vol ou perte d'un conteneur.
- **Force majeure** : Généralement, ce sont des conflits entre le fournisseur et le client à destination.
- **Customs inspection** : Les marchandises peuvent être confisquées pour : des inspections douanières, si elles portent atteinte à un droit de propriété intellectuelle. Les douanes peuvent, après avoir éliminé les éléments contrefaits les vendre aux enchères, conformément à la loi. Lorsqu'il est impossible d'éliminer les éléments contrefaits, ou périmés les douanes détruisent les marchandises.

En effet, le déroulement du diagnostic interne, la cartographie du processus de gestion et leurs descriptions nous ont permis de réaliser le tableau suivant :

Commentaires	Clients concernés	Nombre de conteneurs
Cargo abandoned by customer, awaiting for stripping	38	145
Container cannot be released due to missing OBL	20	60
Container verified in good condition, will be used for export booking	3	5
Customs documentation delays	49	265
Incorrectly updated in RKEM (state action plan in free text field)	3	6
Legal actions started	72	327
Notification Sent to Customer	70	316
Notification to customer failed	3	4
Called Customer	3	17
Force majeure	4	833
Dsp: Lost/ Stolen	2	3
Customs inspection	11	26

FIG. 15 : Les différentes raisons de la non restitution des conteneurs immobiles selon l'extraction de mars 2021.

D'après le tableau précédent, nous constatons que les commentaires se divisent en deux catégories, à savoir l'avancement relatif à chaque client et la raison principale qui a fait que le client n'as pas encore restitué ses conteneurs. Les notifications envoyées aux clients et les actions légaux sont les états d'avancement les plus fréquents avec près de 40% de conteneurs bloqués concernant environ 50% clients. Les principales causes provoquant des retentions courantes se manifestent par la non remise des documents clients nécessaires à la douane portuaire et les problèmes de force majeure qui ont engendré 50% de conteneurs appartenant à 25% de clients distincts.

2.4 Enseignements tirés & problématique

La gestion des conteneurs en souffrance est un maillon très important dans une compagnie maritime. Par conséquent, la bonne performance de ce dernier représente un enjeu majeur pour l'entreprise et lui permettra d'un côté d'augmenter sa profitabilité à travers les gains engendrés par les rotations effectives des conteneurs, et d'un autre côté garantir la disponibilité de ces derniers pour favoriser les échanges mondiaux et éviter les pénuries qui peuvent fragiliser les importations et les exportations.

A l'issue du diagnostic effectué, qui met en relief l'analyse du processus sous ses différentes facettes tout en exposant les facteurs et les raisons ainsi que les actions menées pour tenter d'affiner la situation critique de la retenue des conteneurs. Nous avons défini plusieurs chantiers sur lesquels nous pouvons agir afin d'améliorer le processus.

Dans un premier lieu, nous avons remarqué l'inexistence de catégorisations formalisées relatives aux clients et conteneurs susceptibles d'être consultées lorsqu'un conteneur est retenu ce qui nous pousse à réfléchir à : Comment peut-on prioriser les segments de clients et à quel caractère de conteneurs correspondent ces derniers ?

De ce fait, l'absence de segmentations prédéfinie impacte la prise de décision stratégique et tactique pour aboutir ensuite au questionnement suivant : Quel outil peut-on adapter pour permettre aux décideurs d'effectuer une meilleure prise de décision selon des critères établis, et de façon plus robuste ?

Enfin, l'absence de KPI clairs et des tableaux de bord adaptés pour l'analyse de la performance du processus nous à mener à poser la question suivante : comment peut-on piloter et suivre la performance du processus de conteneurs en souffrance ?

Ces différents questionnements convergent vers une même réflexion plus approfondie sur la façon avec laquelle nous pouvons améliorer le processus de gestion de conteneurs en général. ce qui nous mène à poser la question principale suivante :

” À l'aide de quels outils pourrions-nous optimiser le processus de gestion des conteneurs en souffrance ” ?

Plaçant ainsi l'optimisation du processus de gestion des conteneurs en souffrance comme la problématique majeure de notre étude. Il est impératif de pallier les problèmes liés aux conteneurs en souffrance. Ceci-dit, même si quelques facteurs problématiques sont difficiles à gérer, la compagnie maritime doit essayer d'en minimiser l'immobilisation des conteneurs pour essayer de réduire les pertes qui lui sont liées .

2.5 Synthèse

À travers ce chapitre, nous avons clairement exprimé le besoin de la compagnie (Maersk Algérie) en vue de l'amélioration du processus de gestion des conteneurs en souffrance. Cependant nous nous sommes principalement axés sur la présentation du processus compte tenu des caractéristiques du marché maritime.

Par la suite, nous avons effectué un diagnostic par lequel nous avons cerné le contexte de notre projet. Ceci nous a permis de faire un état des lieux, affiner notre problématique et détecter des pistes d'amélioration permettant d'apporter des solutions relatives à la gestion de ces conteneurs.

Grâce aux enseignements tirés, nous avons constaté que nous pouvons agir dans l'optique d'optimiser le processus de gestion des conteneurs retenus afin de garantir à l'entreprise un gain en rotation des conteneurs et donc augmenter sa rentabilité.

Le chapitre suivant aura pour objectif de présenter les différents concepts et notions théoriques relatifs aux axes d'amélioration retenus.

Chapitre 3

IA, Apprentissage et Tableau de bord

Dans la plupart des cas, les entreprises cherchent toujours à adapter leurs méthodes de gestion aux différents types de clients, collaborateurs ou même conteneurs dans le cas de Maersk,, il est donc nécessaire pour elles d'identifier l'ensemble des segments qui serviront de référence pour l'entreprise dans l'adaptation de ses processus de gestion.

Le machine learning permet à l'aide l'apprentissage non-supervisé de former des groupes d'individus semblables en tenant compte de plusieurs variables par le biais de plusieurs algorithmes comme (KMeans, CAH, DBSCAN etc...).

Dans ce chapitre nous expliquerons ce que c'est que le Machine Learning, les méthodes utilisées pour effectuer le clustering ainsi que les différents indicateurs d'évaluation des modèles d'apprentissage.

Nous allons ensuite introduire les notions liées au pilotage de la performance à travers un tableau de bord qui va offrir à l'entreprise la possibilité d'agir dans un milieu complexe et de prendre des décisions dans l'urgence et l'incertitude.

3.1 Intelligence artificielle

L'intelligence artificielle est définie par Le Robert dictionnaire comme un ensemble de théories et des techniques développant des programmes informatiques complexes capables de simuler certains traits de l'intelligence humaine (raisonnement, apprentissage...).

"Pour comprendre la renaissance actuelle de ce que nous appelons l'« intelligence artificielle », qui remonte clairement aux débuts de l'informatique, nous devons nous intéresser à la convergence de trois tendances : i) le Big Data, ii) le machine learning (apprentissage machine) et iii) le cloud super-computing (calcul de haute performance sur le cloud). En ce sens, l'essor de l'IA est bel et bien une manifestation de la révolution numérique. L'une de ses lois principales, prédites en 1965 par Gordon Moore, cofondateur du fabricant de puces Intel, stipule que la puissance informatique double tous les deux ans en moyenne, à coût constant. Cette croissance exponentielle résulte des prouesses technoscientifiques continues de la miniaturisation. Celle-ci a permis l'avènement de la micro-informatique, puis celui de la nano-informatique, qui ont débouché, avec leur puissance toujours accrue, sur l'apparition des smartphones et de l'Internet des Objets¹" [MIAILHE Nicolas et HODES Cyrus, 2018]

3.1.1 Machine Learning

"La notion qui englobe toute méthode permettant de construire un modèle de la réalité à partir de données, soit en améliorant un modèle partiel ou moins général, soit en créant complètement le modèle ."[Cornuéjols et Miclet, 2003]

L'apprentissage automatique est un sous-domaine de l'intelligence artificielle qui fait référence au développement de méthodes qui permettent à une machine d'évoluer grâce à un processus d'apprentissage, et ainsi de résoudre les problèmes pour lesquels les algorithmes classiques n'offrent pas de solution. En d'autres termes, lorsque nous ne connaissons pas de solution exacte, nous ne pouvons pas écrire de programme

¹Il s'agit de tout objet physique pouvant émettre des données à travers des capteurs.

informatique. L'apprentissage automatique consiste alors à programmer des algorithmes permettant d'apprendre automatiquement de données et d'expériences passées pour proposer des solutions à ces problèmes.[Gacogne, 2015]

Bien qu'il soit actuellement promu par les nouvelles technologies et les nouveaux usages, l'apprentissage automatique n'est pas un domaine de recherche récent. Nous avons trouvé la première définition en 1959, grâce à l'un des pionniers de l'intelligence artificielle, Arthur Samuel, qui a défini l'apprentissage automatique comme un domaine de recherche visant à renforcer l'apprentissage de la machine sans programmation explicite. En 1997, Tom Mitchell de l'Université Carnegie Mellon a fourni une définition plus précise : *"On dit qu'un programme d'ordinateur apprend de l'expérience E par rapport à une certaine classe de tâches T et à la mesure de performance P, si sa performance aux tâches en T, telle que mesurée par P, s'améliore avec l'expérience E"*.

La tâche T représente l'objectif d'un algorithme d'apprentissage, souvent intégré par l'utilisateur sous forme d'exemples historiques. L'utilisateur effectue une mesure de performance à travers plusieurs métriques sélectionnées par lui même en fonction de la situation de recherche et de l'algorithme d'apprentissage utilisé, il s'agit de la mesure de performance P.

L'expérience E représente toutes les données, qui serviront de base à la machine pour construire un modèle automatique.

Le Machine Learning est à la croisée de plusieurs disciplines :

- Les statistiques : pour l'inférence de modèles à partir de données. I Les probabilités : pour modéliser l'aspect aléatoire inhérent aux données et au problème d'apprentissage.
- L'intelligence artificielle : pour étudier les tâches simples de reconnaissance de formes que font les humains (comme la reconnaissance de chiffres par exemple), et parce qu'elle fonde une branche de l'AA dite symbolique qui repose sur la logique et la représentation des connaissances.
- L'optimisation : pour optimiser un critère de performance afin, soit d'estimer des paramètres d'un modèle, soit de déterminer la meilleure décision à prendre étant donnée une instance d'un problème.
- L'informatique : puisqu'il s'agit de programmer des algorithmes et qu'en AA ceux-ci peuvent être de grande complexité et gourmands en termes de ressources de calcul et de mémoire.

La figure suivante illustre le fonctionnement d'un programme ML :

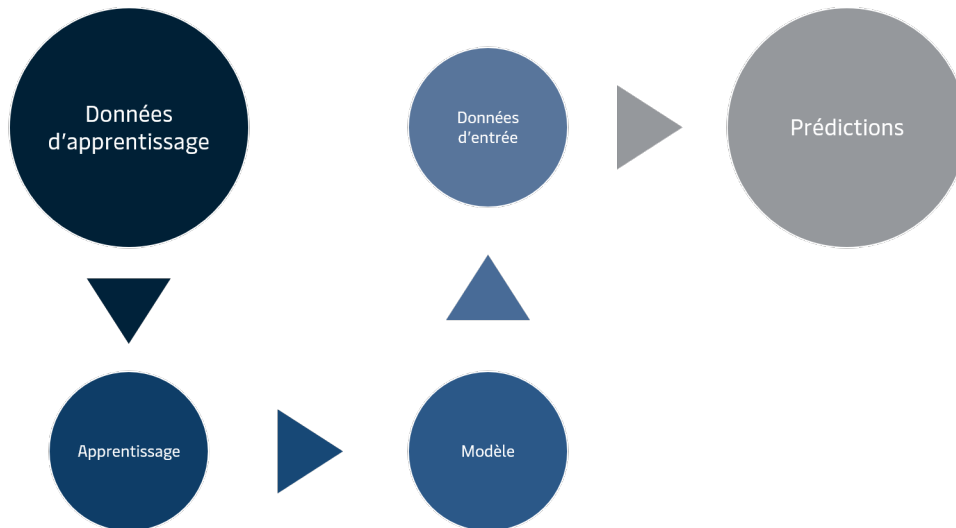


FIG. 16 : Schéma explicatif du fonctionnement d'un programme machine learning

3.2 Types d'apprentissage

Le machine learning nous permet de répondre à des problématiques diverses et variées, voici les 5 types d'apprentissages les plus utilisés en pratique :

- **Supervisé** : On dispose d'un ensemble d'objets et pour chaque objet une valeur cible associée ; il faut apprendre un modèle capable de prédire la bonne valeur cible d'un objet nouveau.
- **Non supervisé** : On dispose d'un ensemble d'objets sans aucune valeur cible associée ; il faut apprendre un modèle capable d'extraire les régularités présentes au sein des objets pour mieux visualiser ou appréhender la structure de l'ensemble des données.
- **Par renforcement** : On dispose d'un ensemble de séquences de décisions (politiques ou stratégiques) dans un environnement dynamique, et pour chaque action de chaque séquence une valeur de récompense (la valeur de récompense de la séquence est alors la somme des valeurs des récompenses des actions qu'elle met en oeuvre) ; il faut apprendre un modèle capable de prédire la meilleure décision à prendre étant donnée un état de l'environnement.
- **Semi-supervisé** : On dispose d'un petit ensemble d'objets avec pour chacun une valeur cible associée et d'un plus grand ensemble d'objets sans valeur cible ; il faut tirer profit à la fois des données avec et sans valeurs cibles pour résoudre des tâches d'apprentissage supervisé ou non supervisé.

- **Actif** : On dispose d'un petit ensemble d'objets avec pour chacun une valeur cible associée ; il faut interagir avec l'utilisateur et lui demander de donner la valeur cible d'un nouvel objet afin de mieux apprendre le modèle de prédiction.

La figure suivante représente les différents types d'algorithmes ML existants ainsi que leurs applications courantes :

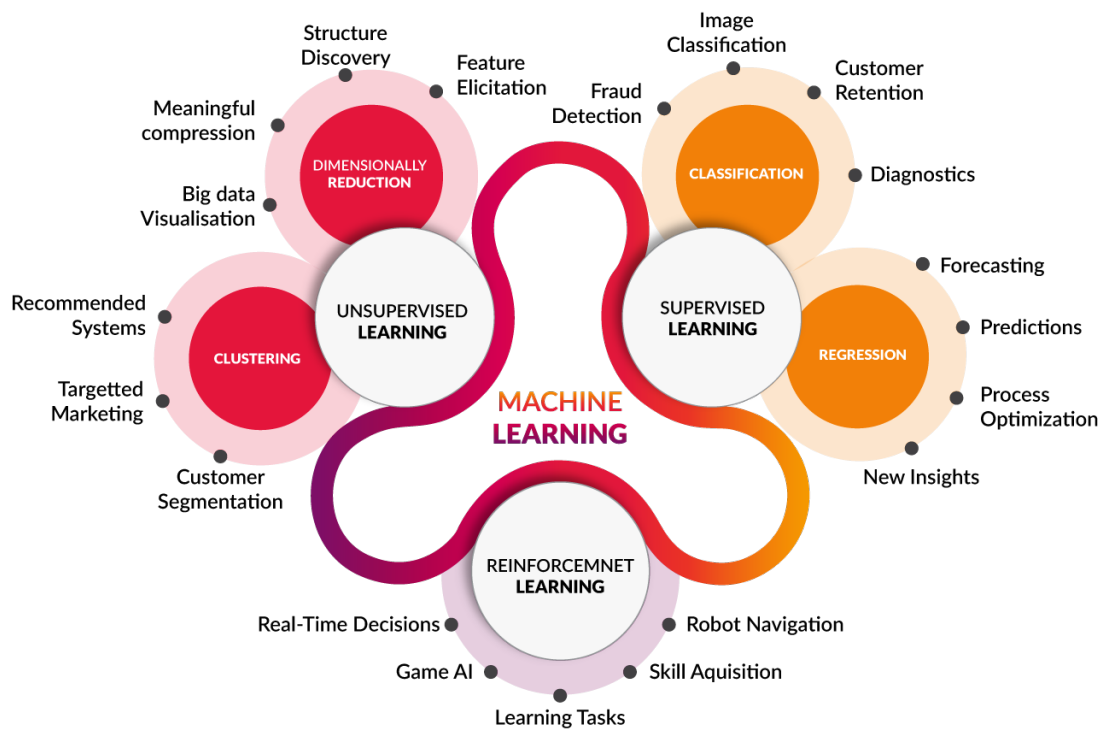


FIG. 17 : Schéma représentant les types d'apprentissage automatique (Kamil Krzyk, 2018)

3.3 Réduction dimensionnelle

"Le nombre de variables prédictives (features) pour un set de données est appelé sa dimension. La réduction de dimensionnalité fait référence aux techniques qui réduisent le nombre de variables dans un ensemble de données, ou encore projettent des données issues d'un espace de grande dimension dans un espace de plus petite dimension." [Khalil Elleuch, 2021]

3.4 Partitionnement de données (clustering)

Le partitionnement de données (clustering, en Anglais) est une technique d'apprentissage non-supervisé englobant plusieurs méthodes de base mathématiques qui

permettent d'identifier des groupes d'objets similaires appelés clusters et cela en fonction des variables qui caractérisent chaque objet.

On note $X = (x_1, x_2, \dots, x_n)$ l'ensemble des n variables caractérisant m objets. Les données sont généralement représentées par une matrice de taille $m \times n$.

Le clustering a donc pour objectif de regrouper les objets jugés similaires dans un même cluster et séparer les objets dissimilaires dans des clusters distincts.

3.5 Les étapes du clustering

Plusieurs étapes sont à considérer lors du développement d'un projet de clustering, en général le processus est composé de trois principales étapes qui sont la préparation des données, le choix de l'algorithme et la validation du modèle de clustering ainsi que son interprétation.

La figure suivante présente le processus de partitionnement de données :

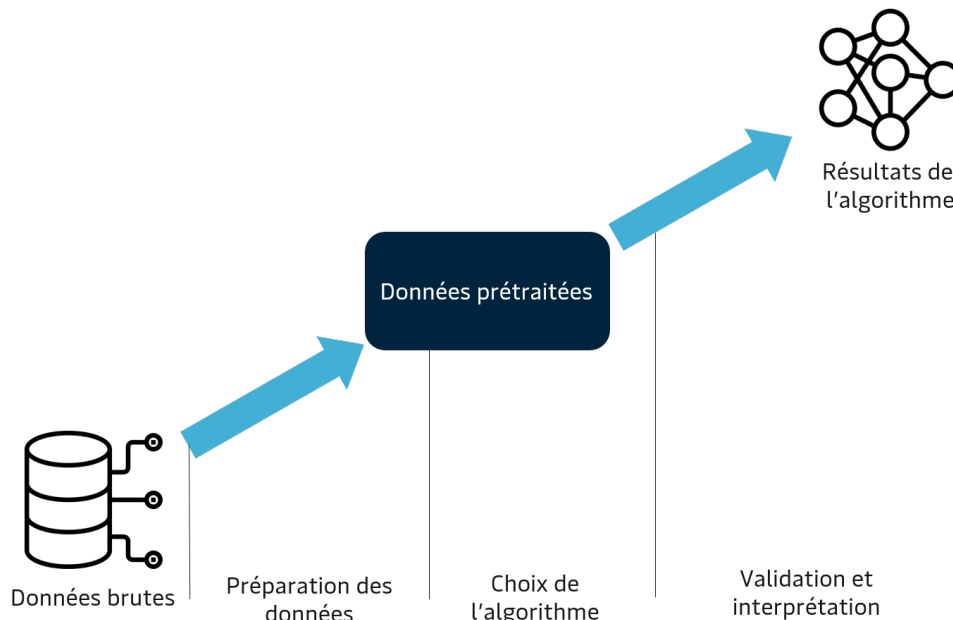


FIG. 18 : Processus de clustering

3.5.1 Préparation des données

Elle fait référence aux opérations de nettoyage et de transformation qui doivent être effectuées sur les données avant de les intégrer dans un modèle d'apprentissage. Il s'agit d'une étape importante qui permet d'introduire à la machine des données fiables et facilement exploitables.

La préparation des données est à son tour un processus composé de plusieurs étapes que nous présentons sur la figure suivante :

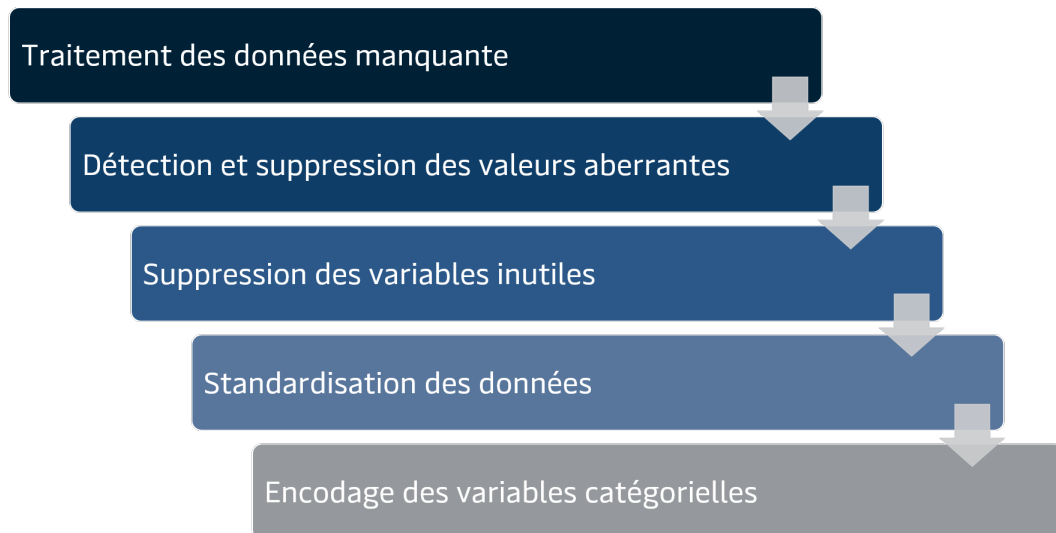


FIG. 19 : Les étapes de préparation de données

3.5.2 Choix de l’algorithme

Il existe une multitude d’algorithmes qui peuvent être utilisés pour un projet de partitionnement de données, l’utilisation de ces algorithmes dépend principalement des données d’entrées. Afin de choisir les algorithmes adéquats, les scientifiques se basent sur des indicateurs qui évaluent la performance des modèles.

Il existe une multitude d’algorithmes qui permettent d’effectuer un partitionnement de données, parmi eux :

K-Means

K-Means est un algorithme d’apprentissage non supervisé développé par le mathématicien Hugo Steinhaus en 1957, il regroupe l’ensemble de données non étiquetées en différents clusters. K définit le nombre de clusters qui doivent être créés dans le processus, si K=2, il y aura deux clusters, et pour K=3, il y aura trois clusters, et ainsi de suite.

Cet algorithme nécessite que le nombre de clusters soit spécifié dès le départ par l’utilisateur. Chaque cluster est associé à des centroïdes qui représentent les moyennes de chaque clusters. l’Algorithme K-means vise à choisir des centroïdes qui minimisent l’inertie, ou le critère de la somme des carrés à l’intérieur du groupe :

$$\sum_{i=0}^n \min_{\mu_j \in C} (||x_i - \mu_j||^2)$$

L’inertie est considérée comme une mesure de cohérence interne au niveau d’un cluster et présente quelques inconvénients :

- L'inertie part du principe que les clusters sont convexes et isotropes, ce qui n'est pas toujours le cas.
- L'inertie n'est pas une métrique normalisée : nous savons simplement que les valeurs les plus faibles sont meilleures et que zéro est optimal. Mais dans les espaces à très haute dimension, les distances euclidiennes ont tendance à être gonflées, dans ce cas l'application d'une ACP pour minimiser les variables d'entrées est préférable afin d'accélérer les calculs.

K-means est souvent appelé l'algorithme de Lloyd. En termes simples, l'algorithme comporte trois étapes. La première étape consiste à choisir les centroïdes initiaux, la méthode la plus simple consistant à choisir des échantillons dans l'ensemble de données . Après l'initialisation, K-means effectue une boucle entre les deux autres étapes. La première étape affecte chaque échantillon à son centroïde le plus proche. La deuxième étape crée de nouveaux centroïdes en prenant la valeur moyenne de tous les échantillons affectés à chaque centroïde précédent. La différence entre les anciens et les nouveaux centroïdes est calculée et l'algorithme répète ces deux dernières étapes jusqu'à ce que cette valeur soit inférieure à un seuil. En d'autres termes, il répète jusqu'à ce que les centroïdes ne se déplacent plus de manière significative.

Minibatch K-Means

Le MiniBatch KMeans est une variante de l'algorithme KMeans qui utilise des petits lots pour réduire le temps de calcul, tout en essayant d'optimiser la même fonction objective. Les mini-batches sont des sous-ensembles de données d'entrée, échantillonnés de manière aléatoire à chaque itération d'apprentissage. Ces mini-batches réduisent considérablement la quantité de calcul nécessaire pour converger vers une solution locale. Contrairement à d'autres algorithmes qui réduisent le temps de convergence des k-means, les mini-batches k-means produisent des résultats qui ne sont généralement que légèrement inférieurs à ceux de l'algorithme standard.

Means Shift

L'algorithme Means shift est un estimateur du gradient de densité non paramétrique développé par Fukunaga et Hostetler en 1975 et exploité récemment par Comaniciu et Meer pour le traitement d'images, la technique means shift est basée sur l'estimateur multidimensionnel de densité.

Étant donné un centroïde candidat x pour une itération t , le candidat est mis à jour en suivant l'équation suivante :

$$x_i^{t+1} = m(x_i^t)$$

Le vecteur means shift est calculé par l'équation suivante :

$$m(x_i) = \frac{\sum_{x_j \in N(x_i)} K(x_j - x_i)x_j}{\sum_{x_j \in N(x_i)} K(x_j - x_i)}$$

Contrairement à l'algorithme k-means et à la classification ascendante hiérarchique, le nombre de cluster est estimé par l'algorithme lui-même par contre l'utilisateur doit introduire la valeur d'un paramètre appelé bandwidth, ce dernier dicte la taille de la région à parcourir par means shift et il peut être estimé à son tour en utilisant un algorithme spécifique. L'algorithme Means shift n'est pas très évolutif, car il effectue plusieurs recherches de voisins les plus proches pendant son exécution. La convergence de l'algorithme est garantie mais il arrête d'itérer lorsque la modification des centroïdes devient très faible. Voici un exemple de clustering en utilisant means shift :

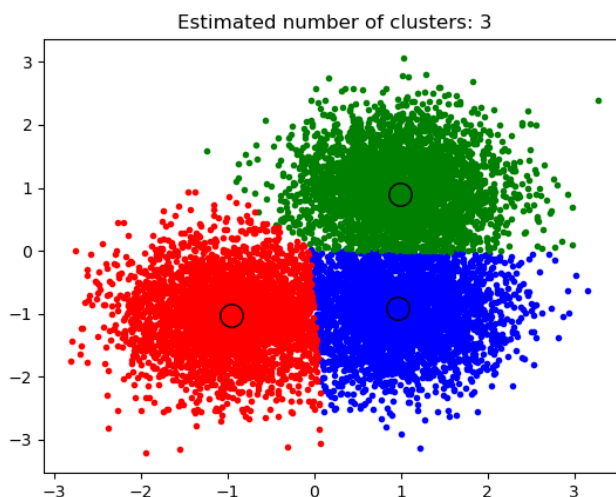


FIG. 20 : Exemple de clustering avec Means Shift
(Documentation Scikit-Learn, 2021)

Classification ascendante hiérarchique

Le clustering hiérarchique est une famille générale d'algorithmes de clustering qui construisent des clusters imbriqués en les fusionnant ou en les divisant successivement. Cette hiérarchie de clusters est représentée sous la forme d'un arbre (ou dendrogramme). La racine de l'arbre est le cluster unique qui regroupe tous les échantillons, les feuilles étant les clusters avec un seul échantillon.

La classification ascendante hiérarchique se base sur un critère de séparation qui doit être choisi par l'utilisateur, après le choix de ce critère la distance euclidienne est calculée :

$$\|a - b\|_2 = \sqrt{\sum_i (a_i - b_i)^2}$$

Cette distance est calculée afin de lier les individus les plus proches entre eux, jusqu'à la génération du dendrogramme en entier.

La figure suivante représente un exemple de dendrogramme :

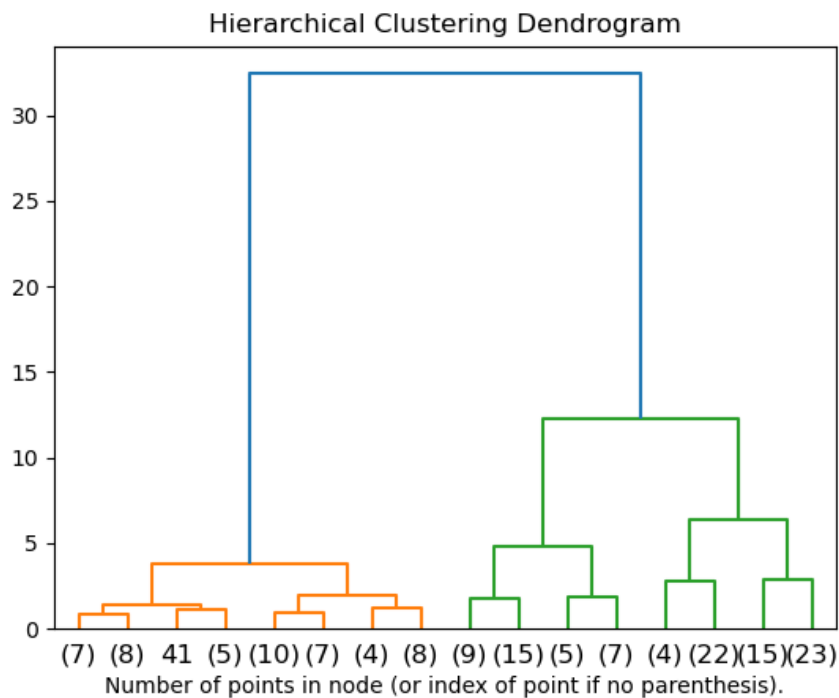


FIG. 21 : Exemple de dendrogramme
(Documentation Scikit-Learn, 2021)

Le dendrogramme est un outil graphique permettant la visualisation du mode de formation des clusters à chaque étape, il permet aussi d'évaluer la similarité des groupes formés.

DBSCAN

C'est un algorithme de clustering basé sur la densité. Il peut découvrir des clusters de différentes formes et tailles à partir d'une grande quantité de données contenant du bruit et des valeurs aberrantes.

L'algorithme DBSCAN utilise deux paramètres :

- **minPts** : Le nombre minimum de points (un seuil) regroupés pour qu'une région soit considérée comme dense.
- **Eps** : Une mesure de distance qui sera utilisée pour localiser les points dans le voisinage de tout point.

DBSCAN est capable d'estimer par lui-même le nombre de clusters nécessaires lors du clustering contrairement à k-means, l'algorithme produit également des résultats plus raisonnables que k-means sur une variété de distributions différentes. La figure ci-dessous illustre ce fait :

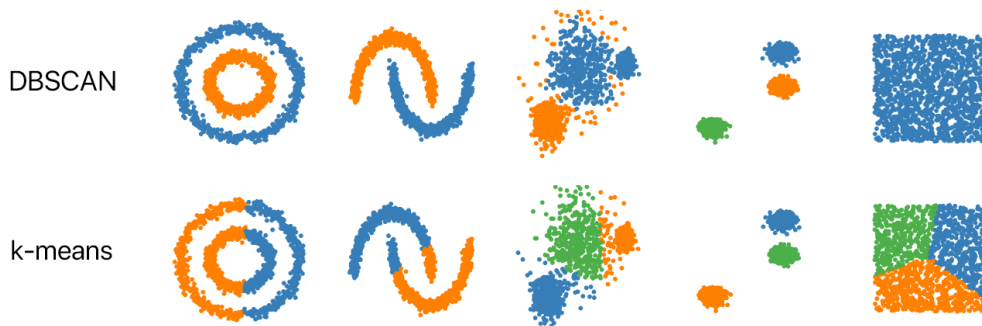


FIG. 22 : Comparaison des résultats obtenus par DBSCAN et K-Means (Documentation Scikit-Learn, 2021)

Au niveau de la figure précédente, nous pouvons remarqués que DBSCAN peut traiter des jeux de données ayant une structure circulaire contrairement au K-Means.

3.5.3 Validation des modèles

En machine learning la validation des données fait référence au processus par lequel les modèles entraînés sont évalués à l'aide de métriques spécifiques à chaque type d'apprentissage.

L'évaluation d'un modèle de clustering s'effectue généralement sur trois différentes étapes illustrées sur la figure suivante :



FIG. 23 : Les étapes d'évaluation d'un modèle de clustering

Mesure de la tendance au partitionnement

Avant d'évaluer la performance du clustering, il est très important de s'assurer que la base de données a une tendance au clustering et ne contient pas de points uniformément distribués. Si les données ne contiennent pas de tendance au clustering, les clusters identifiés par les algorithmes les plus avancés peuvent ne pas être pertinents. La distribution non-uniforme des points dans l'ensemble des données devient importante dans le clustering.

Pour effectuer ce genre d'évaluation, il existe deux principales méthodes :

- Test statistique de Hopkins².
- L'algorithme VAT qui permet de visualiser le nombre de clusters possible pour une base de données précise à travers un graphique.

Détermination du nombre de clusters optimal

Certains algorithmes de clustering, comme les K-means, requièrent le nombre de clusters, k , comme paramètre d'entrée. Obtenir le nombre optimal de clusters est très important dans l'analyse. Si k est trop élevé ou trop faible, alors les points de données seront incorrectement regroupés. Trouver le bon nombre de clusters est un problème très subjectif, cependant il existe quelques méthodes qui permettent d'obtenir le K optimal, il s'agit de :

- Exigences liées au cas étudié ou avis d'experts.
- Approche mathématique en utilisant la méthode du coude.
- Approche statistique par la méthode du gap.

Méthode du coude

La variance intra-groupe est une mesure de la compacité du groupe. Plus la valeur de la variance intra-groupe est faible, plus la capacité du groupe formé est élevée. La somme de la variance intra-classe, W , est calculée pour les analyses de clustering effectuées avec différentes valeurs de k . W est une mesure cumulative de la qualité du clustering des points dans l'analyse. Le tracé des valeurs de k et de la somme correspondante de la variance intra-classe aide à trouver le nombre de clusters.

²Développée par Lawson et Jurs en 1990, la statistique de Hopkins permet de tester le caractère aléatoire spatial d'un jeu de données

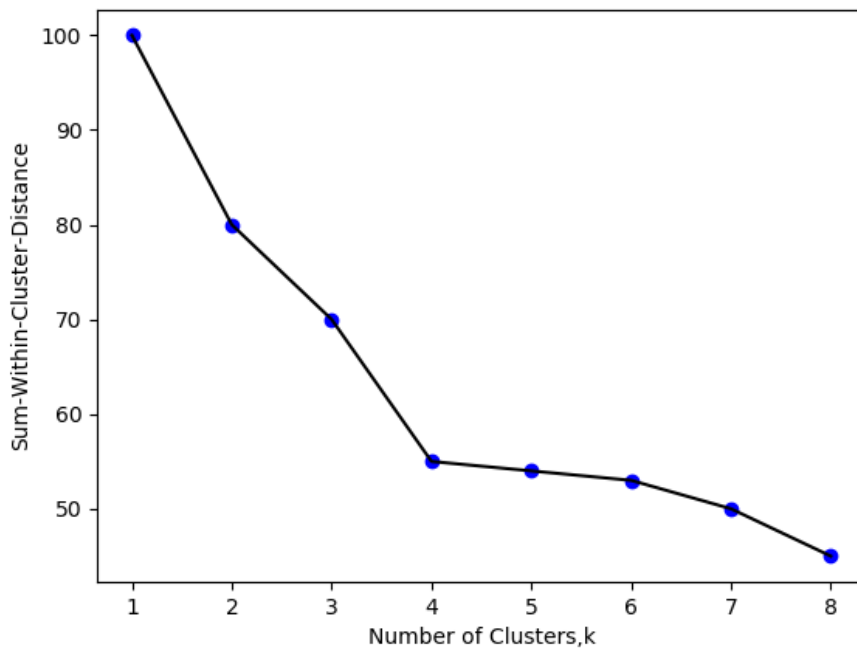


FIG. 24 : Exemple illustrant la méthode du coude

Nous remarquons sur cette exemple que le coude se casse sur $k = 4$, le nombre de cluster optimal dans ce cas est de 4.

La méthode Gap statistics

La méthode a été publiée par R. Tibshirani, G. Walther et T. Hastie (Stanford University, 2001). L'approche peut être appliquée à toute méthode de clustering. La méthode consiste en le calcul des variances intra-clusters pour différentes valeurs de K (nombre de clusters), sa valeur optimale sera la valeur qui maximise les écarts.

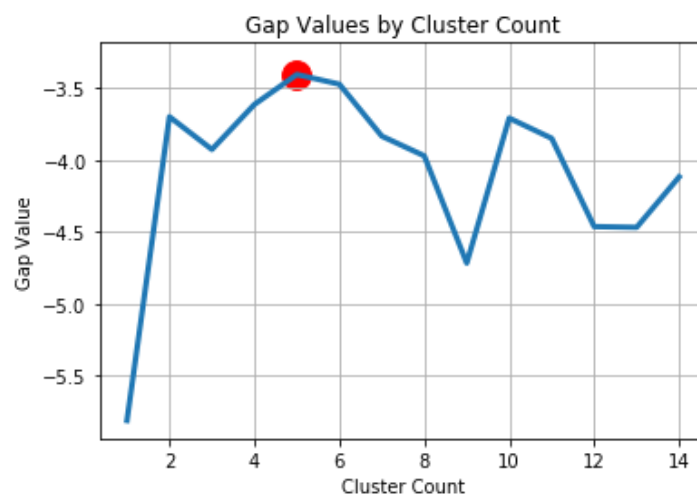


FIG. 25 : Exemple illustrant la méthode gap statistics

Le nombre de cluster optimal est celui qui correspond à la valeur maximale du "Gap Value", dans l'exemple précédent nous pouvons facilement remarquer que la valeur est de 5.

Mesure de la qualité du clustering

Après avoir effectué le clustering il est nécessaire d'évaluer la qualité de ce dernier en utilisant des métriques. Un clustering idéal est caractérisé par une distance intra cluster minimale et une distance inter cluster maximale.

Il existe principalement deux types de mesures qui permettent d'évaluer un modèle de clustering :

- **Mesures extrinsèques** : Ce type de mesure est utilisé dans le cas où le processus de clustering est appliqué sur des données étiquetées, cela veut dire que la nature des clusters est connue en amont du processus. Parmi les mesures utilisées on cite :
 - * Indice de Rand : Il mesure la similarité entre deux partitions d'un ensemble.
 - * Indice Fowlkes – Mallows : Il permet de mesurer la similarité entre deux différents clusters.
 - * Indice d'homogénéité.
 - * Indice d'exhaustivité.
- **Mesures intrinsèques** : Il s'agit de l'ensemble des métriques utilisées pour l'évaluation du clustering effectué sur un ensemble de données non étiqueté. Il existe trois principaux indicateurs qui permettent de mesurer la performance de ce type de clustering :

Coefficient de silhouette :

Il a été publié par Rousseeuw en 1987, c'est une métrique utilisée pour mesurer la qualité d'un partitionnement de données, il est compris entre -1 et 1.

Les résultats du coefficient de silhouette sont interprétés de la manière suivante :

0.71 - 1 : Une structure solide a été trouvée, la qualité du clustering est bonne.

0.52 - 0.7 : Structure raisonnable, la qualité du clustering est moyenne.

0.26 - 0.51 : Faible structure.

< 0.25 : Très faible structure.

Le coefficient de silhouette se constitue de deux autres métriques il est calculé par la manière suivante :

$$s = \frac{b - a}{\max(a, b)}$$

Tel que :

a : Représente la distance moyenne entre un échantillon et tous les autres points de la même classe.

b : Il s'agit de la distance moyenne entre un échantillon et tous les autres points du cluster le plus proche.

La figure suivante illustre les deux distances a et b sur un exemple de clustering

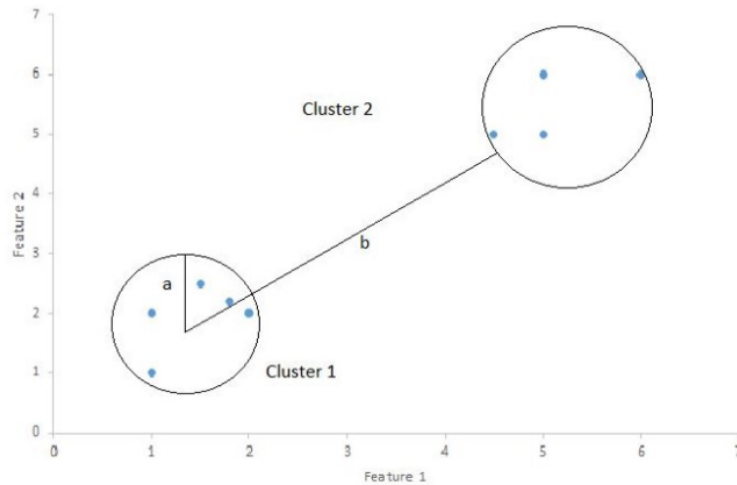


FIG. 26 : Les distances a et b sur un exemple de clustering

Indice de Calinski-harabasz :

Introduit par Calinski et Harabasz en 1974, il s'agit du rapport entre la variance inter-groupes et la variance intra-groupe. La valeur de l'indice est plus élevée lorsque les clusters sont denses et bien séparés, ce qui correspond au concept de cluster. La formule mathématique de l'indice est la suivante :

$$s = \frac{\text{tr}(B_k)}{\text{tr}(W_k)} \times \frac{n_E - k}{k - 1}$$

Tel que :

s : Indice de Calinski-Harabasz.

k : Nombre de clusters.

B_k : Matrice des variances intra-groupes avec :

$$B_k = \sum_q n_q (c_q - c)(c_q - c)^T$$

W_q : Matrice des variances inter-groupes avec :

$$W_k = \sum_{q=1}^k \sum_{x \in C_q} (x - c_q)(x - c_q)^T$$

C_q : L'ensemble des points dans un cluster q .

c_q : Le centre d'un cluster.

c : Le centre de l'ensemble de données E .

Indice de Davies-Bouldin : C'est une métrique utilisée pour la mesure de la qualité d'un groupe de données, il s'agit de la moyenne du rapport maximal entre la distance d'un point au centre du cluster et la distance entre deux centres de cluster.

L'indice est calculé à partir d'un coefficient de similarité :

$$R_{ij} = \frac{s_i + s_j}{d_{ij}}$$

Avec :

R_{ij} : Coefficient de similarité.

s_i : La distance moyenne entre chaque point et son centroïde.

d_{ij} : Distance entre les deux centroïdes.

L'indice est donc défini par la relation suivante :

$$DB = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} R_{ij}$$

Avec k : Le nombre de clusters.

3.5.4 Présentation du langage de programmation python

Python est un langage de programmation le plus utilisé dans le domaine du big data et du machine learning, il dispose d'une multitude de bibliothèques et de méthodes prédéfinies qui permettent d'élaborer des programmes de machine learning et de traitement de données d'une manière très efficace, parmi les utilisations du langage de programmation python, nous citons :

- création des bots.
- Scarping de sites web.
- Analyse et visualisation de données.
- Apprentissage automatique.

- Développement web à l'aide des frameworks Django et Flask.
- Développement de jeux à travers Pygame.
- Développement d'application mobile en utilisant des frameworks comme Kivy.

Selon Google trends python est aujourd'hui le langage de programmation le plus recherché et le plus populaire au monde :

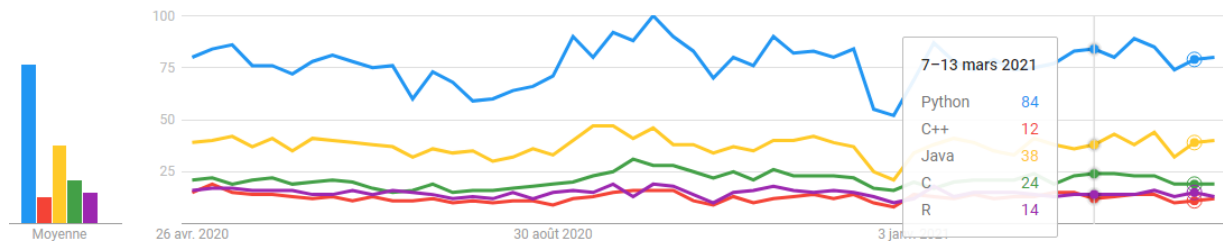


FIG. 27 : Capture d'écran, Google trends

3.5.5 Bibliothèques python au service du machine learning

Une bibliothèque python est un ensemble de Fonctions, classes, objets, constantes etc. prédéfinis ajoutant plus de possibilités aux utilisateurs dans la manipulation des données, parmi ces bibliothèques nous citons celles qui sont les plus utilisés dans la machine learning :

- **Numpy** : Conçue afin de gérer les matrices et tableaux multidimensionnels ainsi que les différentes fonctions qui s'exécutent sur ces tableaux. Numpy permet aussi de créer et d'enregistrer des tableaux directement sur des fichiers, manipuler des vecteurs, matrices et polynômes.
- **Pandas** : Bibliothèque python open source, elle permet d'effectuer des calculs sur des tableaux de données numérique et sur des séries temporelles et cela à travers des méthodes mathématiques et statistiques prédéfinies sur la bibliothèque.
- **Matplotlib** : Matplotlib est une bibliothèque utilisée pour tracer et visualiser des données sous forme de graphiques. Il peut être utilisé en conjonction avec les bibliothèques de calcul scientifique Python comme NumPy.
- **Scikit-learn** : C'est une bibliothèque contenant une multitude d'algorithmes d'apprentissage pré-définies qui ont été développés par de nombreuses personnes issues du monde académique.
- **Seaborn** : Elle permet de visualiser des données à travers des graphiques pré-implémentés, seaborn a été développé en se basant sur la bibliothèque matplotlib.

3.6 Performance et tableaux de bords

Dans la partie suivante , nous nous intéresserons dans un premier temps à la notion de la performance, dans un second temps nous détaillerons les concepts liés aux tableaux de bord.

3.6.1 La performance

Nous allons délimiter, dans cette partie, les contours de la notion de performance et d'indicateurs de performance :

Définition de la performance : La performance reflète la capacité de l'entreprise à construire de la valeur à court, moyen et long terme, en se basant sur l'analyse des résultats passés. Elle désigne donc le degré d'accomplissement d'objectifs préalablement établis par le comité responsable de l'organisation.

Selon G.Charreaux (1998) , “la performance est la capacité de l'entreprise à concrétiser ses objectifs stratégiques en adoptant les meilleures façons de faire” [2] G. CHARREAUX, Le point sur la mesure de performance des entreprises, Paris, Ed. Economica, 1998, p7 Cette notion peut être représentée par le modèle de Gilbert et son triangle de la performance suivant :

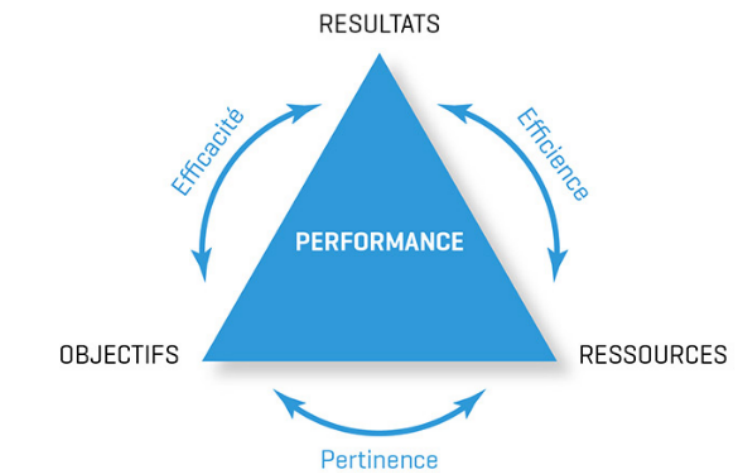


FIG. 28 : La performance cercle vertueux entre résultats et moyens consacrés.

3.6.2 Tableaux de bord

Nous visons à travers cette partie à faire ressortir les concepts liés aux TDB , leurs typologies ainsi que les démarches adoptées pour concevoir ces derniers.

Définitions et caractéristiques

Un tableau de bord, selon Boix et Féminier (2003), est un outil destiné au responsable pour lui permettre grâce à des indicateurs, présentés de manière synthétisée, de contrôler le fonctionnement de son système en analysant les écarts significatifs afin de prévoir et de décider pour agir.

“C’est un outil d’aide à la décision” (Selon Selmer, 1998) primordial pour toute entreprise voulant se distinguer des autres sur le marché.

Types de tableaux de bord

La finalité principale d’un TB est de piloter une activité, c’est-à-dire prendre des décisions pour agir à partir de résultats constatés. Il peut toutefois également être utilisé pour analyser une situation, comprendre les raisons d’échec ou succès, anticiper des menaces, etc. On distingue donc 3 finalités qu’on peut citer par ordre d’importance selon le schéma de “Igor Ansoff” :

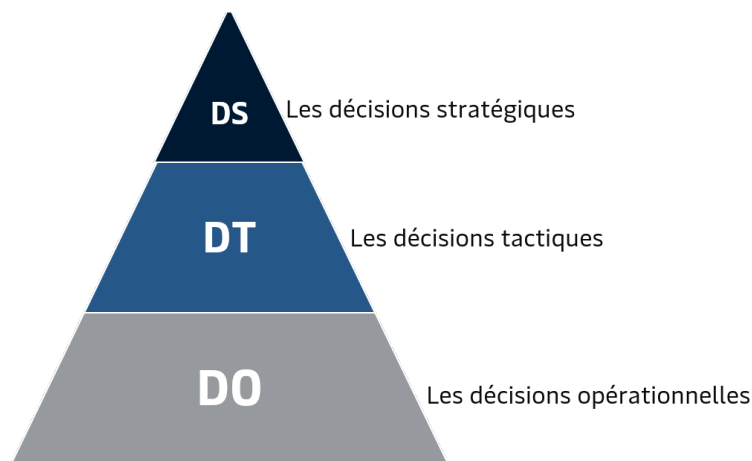


FIG. 29 : La pyramide d’Ansoff : les trois types de décisions en entreprise.

- **TDB Stratégiques** : aussi appelé TDB prospectif ”balanced scorecard”, qui est axé, comme son nom l’indique, sur la stratégie de l’entreprise, permettant de piloter la performance en s’appuyant sur les facteurs clés de succès de l’activité. C’est un outil de pilotage à long terme élaboré généralement à une fréquence annuelle ou semestrielle.
- **TDB Tactique** : également appelé TDB de gestion, engage l’entreprise à moyen terme à fin d’évaluer le niveau de performance des processus pilotés régulièrement à une fréquence hebdomadaire ou mensuelle.
- **TDB Opérationnel** : C’est un outil de pilotage d’un horizon essentiellement à court terme généré d’une manière journalière ou hebdomadaire. Son objectif comprend deux volets : la mesure de l’avancement des plans d’actions décidées et l’évaluation de leur performance.

Les indicateurs de performance

Selon Voyer, un indicateur est *”un élément ou un ensemble d’éléments d’information significative, un indice représentatif, une statistique ciblée et conceptualisée selon une préoccupation de mesure. Il résulte de la collecte de données sur un état, sur la manifestation observable d’un phénomène ou sur un élément lié au fonctionnement d’une organisation ”*.

Les indicateurs de performance ou “ KPI : Key Performance Indicators” sont des données quantifiées permettant de mesurer l’efficacité et/ou l’efficacité d’un processus , projet ou de l’activité de l’entreprise au vu des objectifs fixés .

Dans ce sens et d’une manière plus simple : ce sont des chiffres , ratios , moyennes ou autres qui ont pour vocation le contrôle et l’amélioration en continu de l’organisation.

3.6.3 Méthodologies de construction d’un tableau de bord

L’élaboration d’un tableau de bord est un processus qui diffère selon la méthodologie choisie, qui doit être parfaitement adapté à la stratégie et aux besoins de l’entreprise.

La littérature managériale propose différentes méthodes pour concevoir et construire ces outils de pilotage parmi lesquelles :

Méthode GIMSI : *”GIMSI est une méthode de conception du système de pilotage, point central du Business Performance Manager.*

Structuré en 10 étapes, la méthode s’inscrit naturellement dans un mode de management moderne fondé sur un principe de gouvernance généralisé privilégiant la prise de décision répartie. Multiplier les points de décision, rapprocher le processus décisionnel au plus près du terrain, là où se situe l’information, là où l’action est possible, est en effet l’unique moyen de maîtriser la complexité croissante des organisations.

”Les décideurs ne sont pas isolés. la méthode GIMSi favorise la coopération entre les décideurs, le partage de la connaissance et l’intégration performante des outils et techniques de l’implémentation des tableaux de bord.” [Alain Fernandez, 2003]

La méthode GIMSI est structurée en 4 phases, chaque phase comporte plusieurs étapes (10 au total), chacune des 10 étapes marque un seuil identifiable dans l’avancement du système :

L'enchaînement des étapes est présenté au niveau de la figure suivante :

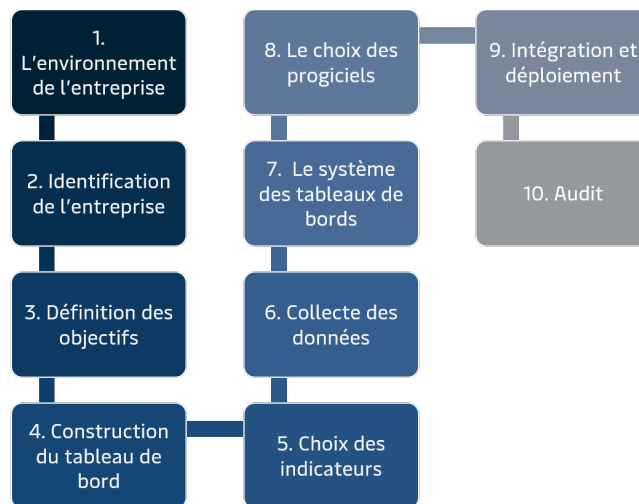


FIG. 30 : Les étapes de la méthode GIMSI

- **L'environnement de l'entreprise** : Analyse de l'environnement économique et de la stratégie de l'organisation afin de définir le périmètre et la portée du projet.
- **Identification de l'entreprise** : Analyse des structures de l'organisation pour identifier les processus, activités et acteurs concernés.
- **Définition des objectifs** : Sélection des objectifs tactiques de chaque équipe et des indicateurs.
- **Construction du tableau de bord** : Définition du tableau de bord de chaque équipe.
- **Choix des indicateurs** : Choix des indicateurs en fonction des objectifs choisis.
- **Collecte des données** : Identification des informations nécessaires à la construction des indicateurs.
- **Système des tableau de bords** : Construction du système de tableaux de bords et vérification de la cohérence globale.
- **Choix des progiciels** : Elaboration de la grille de sélection pour le choix des progiciels et outils adéquats.
- **Intégration et déploiement** : Implantation des progiciels, déploiement à l'organisation.
- **Audit du système** : Suivi permanent du système.

Méthode OVAR : OVAR "Objectifs - Variables d'Action - Responsabilité" est un outil de pilotage mis au point par 3 enseignants du groupe HEC. À l'instar du tableau de bord prospectif, en s'appuyant sur une méthode formalisée sous forme de processus, la méthode OVAR a pour vocation d'aider les dirigeants à élaborer et déployer une stratégie à tous les niveaux. Avec une attention particulière portée à la cohérence entre l'ensemble des domaines. Le déroulement de la méthode OVAR est illustré par la figure suivante :

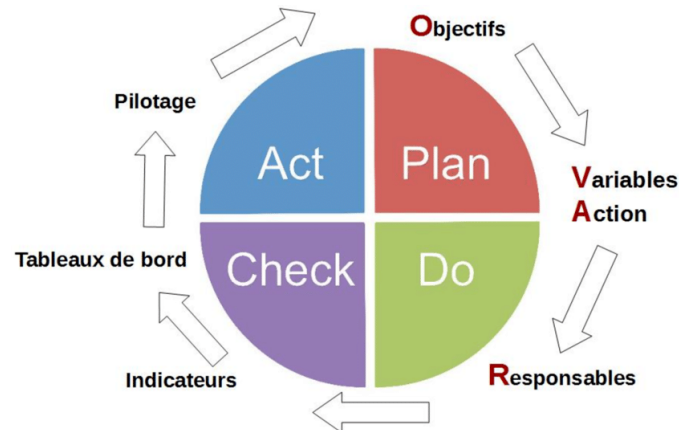


FIG. 31 : Le déroulement de la méthode OVAR.

Méthode de Balanced Scorecards "BSC" (Tableau de bord prospectif) : Elle a été élaborée en 1990 par Robert Kaplan et David Norton à la Harvard Business School partant du principe que la performance de l'entreprise ne se résume pas à la performance financière d'où son nom s'inspire. En effet le BSC a pour vocation de trouver l'équilibre entre les indicateurs financiers et non-financiers. Ce modèle de pilotage émerge de la stratégie de l'entreprise, celle-ci se voit entourée de quatre axes comme le montre la figure qui suit :

3.6.4 Technologies de visualisation :

La visualisation des données semble être une étape importante, quelles que soient les données qu'on veut analyser. Cependant il existe plusieurs outils pour obtenir une visualisation propre et attrayante, voici notre revue de quelques outils utilisés pour l'analyse des données :



FIG. 32 : Figure illustrant les différents outils informatiques de visualisations.

- **Tableau** : Grâce à sa simplicité d'utilisation et à sa capacité à produire des visualisations interactives, ce logiciel compte une très large base de clients. Il permet de : créer des graphiques, des cartes et de nombreux autres graphiques à travers une version gratuite appelée Tableau Public et une version payante.

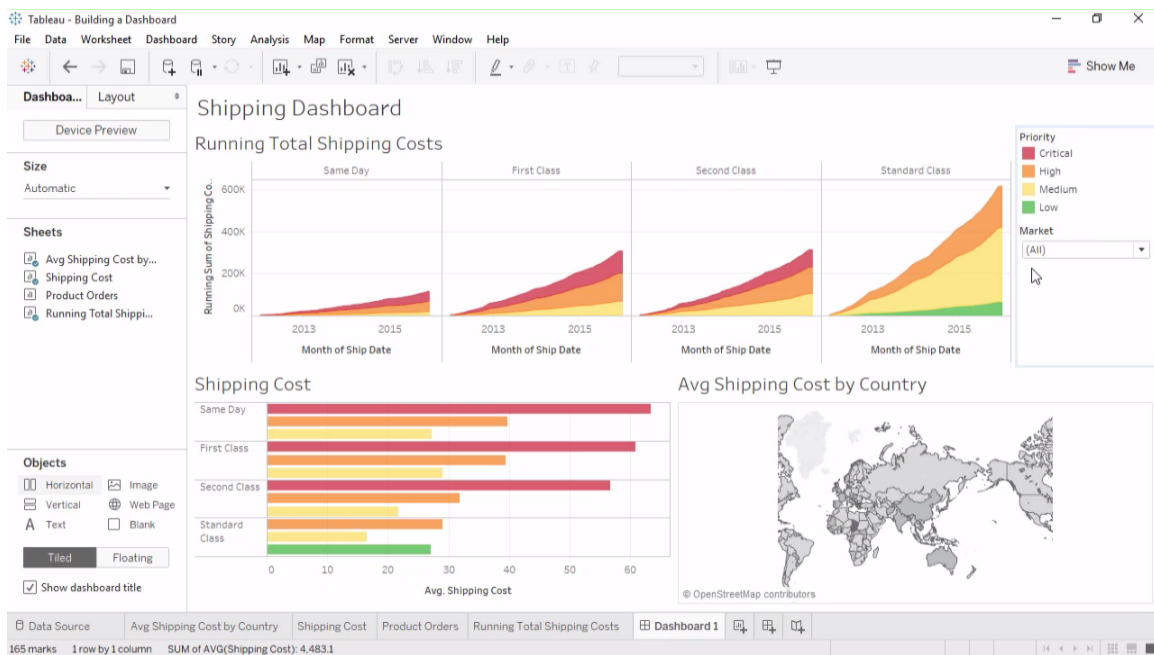


FIG. 33 : Interface Graphique d'un TDB sous Tableau .

- **Zoho Analytics** : il s'agit d'un logiciel de BI , d'analyse de données et de gestion de projets Il permet : d'analyser les données et d'avoir une synchronisation quotidienne des rapports et des KPI.

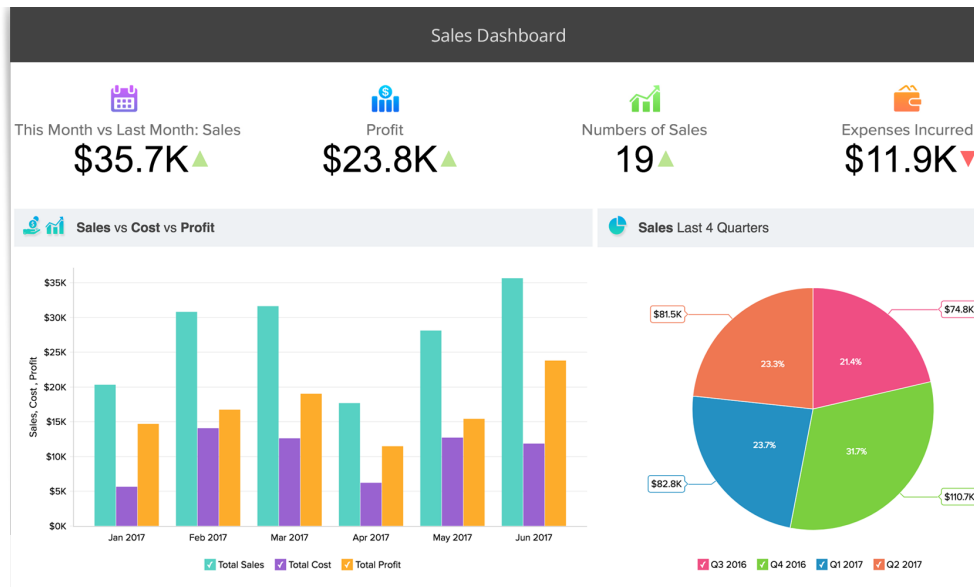


FIG. 34 : Interface Graphique d'un TDB sous Zoho Analytics .

- **Microsoft Power BI :** Il s'agit d'une solution de business intelligence développée par Microsoft en 2014, elle permet aux entreprises d'agréger, d'analyser et de visualiser des données provenant de plusieurs sources. Il permet de :
 - Se connecter aux données, où qu'elles soient. Explorer ensuite ces données avec d'incroyables visualisations interactives.
 - Publier des rapports et des tableaux de bord, collaborer avec toute l'équipe et partager des insights à l'intérieur et à l'extérieur de l'organisation.

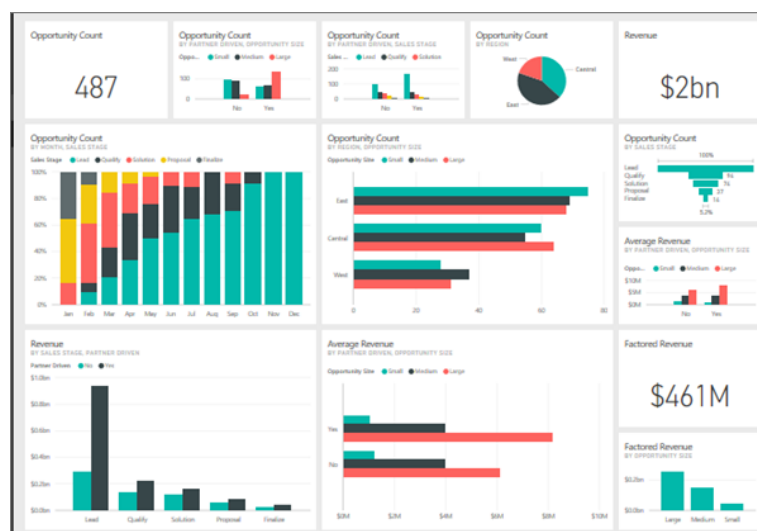


FIG. 35 : Interface Graphique d'un TDB sous Microsoft Power BI.

- **Sisence** : Créée en 2004, Sisence est une plateforme BI disponible en ligne permettant de synthétiser des données complexes et de les afficher sous forme de tableaux de bord. Il permet de :
 - Créer des tableaux de bord interactifs sans compétences techniques et les partager.
 - Offrir aux utilisateurs la liberté d'interroger les données en temps réel.

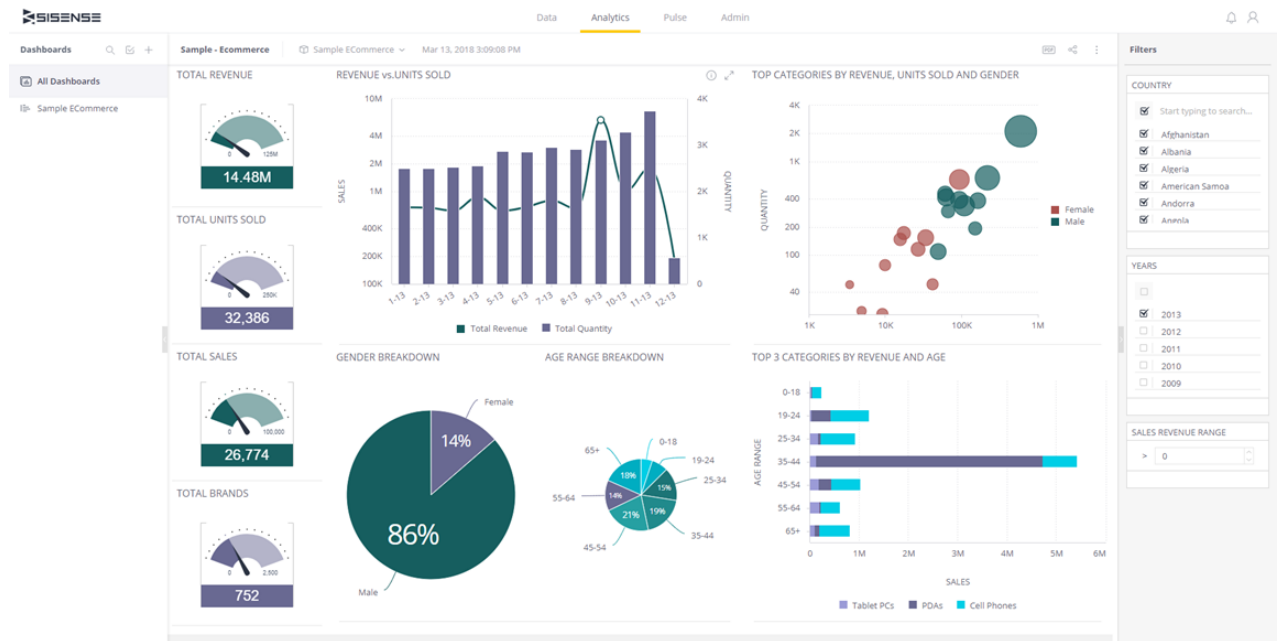


FIG. 36 : Interface Graphique d'un TDB sous SISENCE.

3.7 Synthèse

Ce chapitre, en plus d'introduire le domaine du machine learning, nous a permis d'explicitier les étapes de la démarche du clustering, car nous avons décidé de nous en servir comme base de notre méthodologie dans le déploiement de notre solution. Il nous a aussi permis de faire le tour des algorithmes d'apprentissage non-supervisé en expliquant leurs fonctionnements afin de faciliter leurs compréhensions. Nous avons cloturé ce chapitre par l'explication des méthodes utilisé dans l'élaboration des tableaux de bord ainsi que les progiciels et outils informatiques associés.

Chapitre 4

Segmentations, clustering et Tableau de bord

Suite au diagnostic que nous avons élaboré dans chapitre 2 nous avons pu repérer plusieurs pistes d'améliorations possibles au niveau des outils de gestion des conteneurs en souffrance de Maersk. Nous présenterons dans ce chapitre l'ensemble des préconisations que nous avons effectuées ainsi que les différentes étapes de leurs mise en place.

Afin d'assurer un gain de temps dans la gestion des retours, il est nécessaire de bien comprendre les différents cas qui se présente à l'entreprise, en général un cas est composé d'un clients et d'un ou plusieurs conteneurs, nous avons donc opter pour une double segmentation (Clients et conteneurs) dans le but de s'assurer que l'ensemble des paramètres soient pris en considération.

Cette segmentation permettra au personnel de Maersk de :

- Amliorer la réactivité face aux différents cas de rétention de conteneurs.
- Adapter les méthodes de gestion et effectuer des actions diférenciées en fonction de chaque cas.
- Enregistrer un gain de temps en précisant en amont les stratégies liées à la gestion des conteneurs en souffrance.

Nous avons donc opter pour le clustering pour pouvoir regrouper l'ensemble des observations contenant des caractéristiques similaires en plusieurs groupe appelé clusters et cela en se basant sur des variables qui seront sélectionnés par la suite.

Pour l'implémentation de la solution, nous avons choisi le langage python car il contient la totalité des fonctions dont nous aurons besoin pour effectuer un clustering, et c'est un langage qui est assez simple à comprendre pour un ingénieur qui n'a pas des connaissances poussée en programmation informatique.

Nous avons choisi Google Colaboratory comme environnement de développement car il nous a permis de travailler en mode collaboratif à travers le web, il est défini comme suit :

"Google Colab ou Colaboratory est un service cloud, offert par Google (gratuit), basé sur Jupyter Notebook¹ et destiné à la formation et à la recherche dans l'apprentissage automatique. Cette plateforme permet d'entraîner des modèles de Machine Learning directement dans le cloud. Sans donc avoir besoin d'installer quoi que ce soit sur notre ordinateur à l'exception d'un navigateur."[Henri Michel, 2019]

La figure suivante illustre le plan de notre projet, tel que le premier mois du stage a été réservé au diagnostic de la problématique et les deux mois qui suivent pour la mise en place de la solution que nous avons proposé :

¹C'est un projet et une communauté dont le but est de développer des logiciels open source

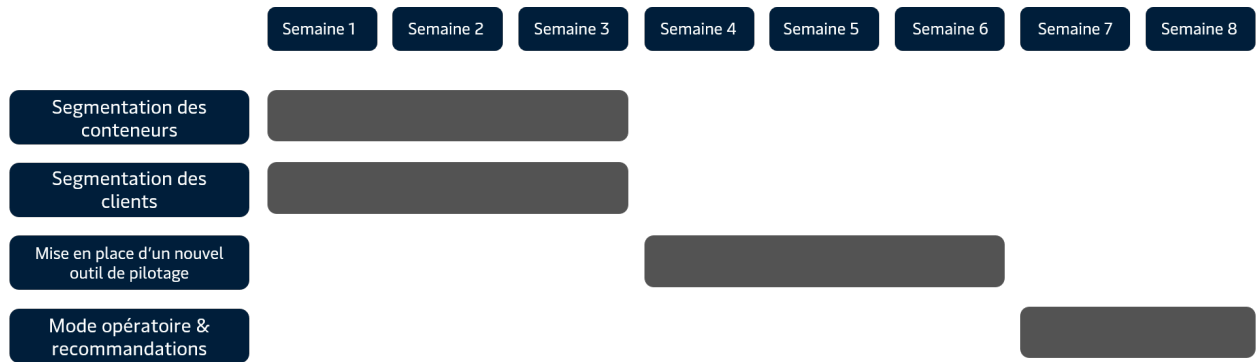


FIG. 37 : Diagramme de GANTT du projet

4.1 La démarche CRISP-DM

Nous avons opté pour l'apprentissage non supervisé et les algorithmes de clustering pour segmenter les clients et les conteneurs, la démarche la plus adaptée pour ce type de projet est CRISP-DM.

Il s'agit d'une méthode de travail développée en 1960 par IBM et reste aujourd'hui la seule méthode qui permet d'orienter les projets d'exploration des données, elle est très utilisée dans les projets de Machine learning et big data.

La figure suivante illustre la méthodologie CRISP - DM combinée avec le processus de clustering :



FIG. 38 : Schéma des étapes de la démarche CRISP-DM

4.2 Segmentation des conteneurs

Au niveau de cette partie, nous allons déroulés la démarche CRISP-DM combinée avec le processus de clustering afin de répartir l'ensemble des conteneurs ayant des caractéristiques similaires dans un même groupe.

4.2.1 Compréhension du cas

Cette première étape de la méthodologie CRISP-DM a pour but de contextualiser notre étude par rapport à l'objectif visé, celle-ci concerne dans notre cas le développement d'un nouvel outil d'aide à la décision qui permettra à Maersk de faciliter le traitement des cas de conteneurs en souffrance et cela commence par la résolution d'un problème de

clustering sur ces même conteneurs.

Pour mieux détailler cette première partie, nous pouvons citer les points suivants :

- La base de donnée brute a été obtenue en effectuant une extraction d'un fichier .xls à partir du système de tracking des conteneurs de Maersk, elle contient 5590 lignes et 15 variables, chaque ligne correspond à un conteneurs, on peut distinguer deux types de variables dans notre base de données il s'agit des variables numériques et des variables catégorielle que nous allons encoder par la suite.
- La base de données ne contient donc pas de variable cible, nous essayerons de regrouper uniquement les conteneurs qui ont des caractéristiques similaires.
- Les données dont nous disposons ne sont pas étiquetées, les clusters ne sont pas précisés en amont.
- Nous avons opté pour le langage python pour implémenter les algorithmes de clustering à l'aide des bibliothèques Pandas, Numpy, Matplotlib, Seaborn et Scikit-learn.

4.2.2 Compréhension des données

Cette étape va nous permettre de bien explorer la base de données brute et de mieux la comprendre afin de pouvoir tirer les informations les plus pertinentes au profit de notre étude.

Nous pouvons visualiser des informations générales sur la base de données en utilisation la fonction `info()` de pandas :

```
0    Days          5569 non-null  int64
1    Equipment Number  5569 non-null  object
2    Eqp Size        5569 non-null  int64
3    Last Loc Code   5569 non-null  object
4    Last Loc Name   5569 non-null  object
5    Last Move       5569 non-null  object
6    1st TPDoc       5569 non-null  object
7    TPDoc Pkg Ct    5565 non-null  float64
8    TPDoc Pkg Type  5553 non-null  object
9    H a z           9 non-null    object
10   IMO Class       9 non-null    float64
11   UNNO            9 non-null    float64
12   Marine Pollutant  9 non-null    object
13   O O G           32 non-null    object
14   FRT RLS         5266 non-null  object
15   *Standard Comment  5569 non-null  object
16   Eqp Type        5569 non-null  object
17   Final Vsl Name  5569 non-null  object
18   Cargo wgt kgs   5569 non-null  int64
19   Business Unit   5569 non-null  object
20   Shipper Customer  5569 non-null  object
21   Consignee Customer  5568 non-null  object
dtypes: float64(3), int64(3), object(16)
```

FIG. 39 : Visualisation des variables et leurs types sur python

La base de données contient au total 6 variables numériques, trois de type float² et trois

²Varibale contenant des nombre réels.

de type int³ ainsi que 16 autres variables catégorielles, la figure suivante représente la répartition des types de variables :

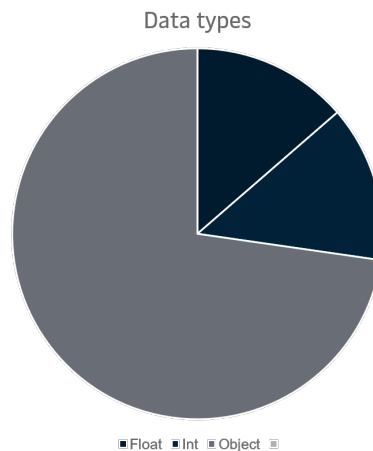


FIG. 40 : Diagramme en camembert représentant la répartition des types de variables

Days : Représente le nombre de jours qu'un conteneur est retenu, elle est comptabilisée à partir de la date d'accostage du navire au niveau du port d'arrivée.

Equipment Number : Il s'agit du numéro d'identification du conteneur, il est donc unique pour chaque conteneur.

Eqp Size : Elle fait référence à la taille du conteneur et peut prendre trois valeurs différentes : 20, 40 ou 45 pieds.

Last Loc Code : Cette variable représente le code du port d'arrivée, par exemple : Le code du port d'Alger est DZALGTM, les deux premières lettres font référence au pays.

Last Loc Name : Il s'agit du nom complet du port.

Last move : Elle indique si le conteneur a été transféré du port vers sa destination finale.

1st TPDoc : Elle représente le numéro du connaissance⁴.

TPDoc Pkg Type : Il s'agit du type de packaging pour lequel le client a opté pour charger sa marchandise.

TPDoc Pkg ct : Elle indique la quantité de package au niveau de chaque conteneur.

UNNO : Cette variable représente le numéro mis en place par les nations unies afin d'identifier le type de produits dangereux présents dans le conteneur.

H a z : Cette variable nous indique si le conteneur contient des produits dangereux ou non.

IMO Class : Il s'agit de la variable qui précise la classe du produit dangereux s'il existe, il peut y avoir 9 différentes classes de produits dangereux selon la catégorisation de Maersk.

Marine Pollutant : Elle fait référence aux produits polluants.

O O G : Il s'agit des marchandises spéciales comme les marchandises hors gabarit.

FRT RLS : Elle indique si le client s'est rendu au niveau de l'agence afin de récupérer les documents nécessaires pour libérer sa marchandise.

Final Vsl Name : Représente le nom du navire.

³Variable contenant des nombres entiers.

⁴Le connaissance est le document matérialisant le contrat de transport maritime conclu entre le chargeur et le transporteur maritime.

Cargo Wgt Kgs : Il s’agit du poids de la marchandise en kilogramme.

Eqp type : C’est le type du conteneur, la variable peut prendre quatre différentes valeurs : Dry, Reefer, Open et Flat.

Shipper Customer : Elle représente le numéro d’identification du client expéditeur de la marchandise.

Consignee Customer : Il s’agit du numéro d’identification du client final.

Business unit : C’est l’entité responsable de l’exportation de l’origine, lieu de provenance du conteneur.

Standard comment : C’est les commentaires représentant l’état actuel ou les raisons pour lesquelles un conteneur n’a pas été restitué par le client.

Valeurs manquantes :

”Souvent les bases de données brutes contiennent des valeurs manquantes, ces dernières doivent être détectées et traitées pour ressortir des données de meilleure qualité et poursuivre le processus avec une base de données complète. Les données manquantes ont des origines matérielles diverses. Des valeurs peuvent être absentes soit parce qu’elles n’ont pas été observées, ou qu’elles ont été perdues ou étaient incohérentes” [Glasson-Cicognani et Berchtold, 2010].

En premier lieu nous avons supprimé les doublons par la fonction `drop_duplicate()` de pandas qui utilise comme attribut une variable distinctive, dans notre cas c’était le numéro d’identification du conteneur “Equipment Number”.

Par la suite nous avons pu détecter les variables contenant des valeurs manquantes par une `Missing map`⁵ obtenue à l’aide de la fonction `heatmap()` de la bibliothèque `seaborn` :

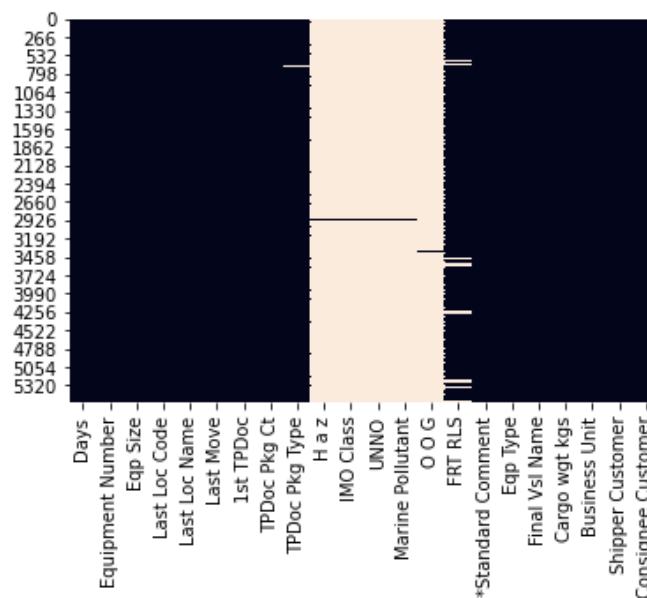


FIG. 41 : Heatmap montrant les valeurs manquantes.

⁵Représentation graphique permettant la visualisation des valeurs manquantes

A l'aide du graphique précédent, nous pouvons facilement voir que les données manquantes se trouvent au niveau des variables (H a z, IMO Class, Marine Pollutant, OOG, FRT RLS, TPDoc Pkg Type).

Valeurs aberrantes et anomalies : *”La valeur ou l’observation aberrante est statistiquement discordante dans le contexte d’un modèle de probabilité supposé connu”* [Barnett et Lewis, 1994; Grubbs, 1969; Munoz-Garcia et al., 1990; Pires et Santos-Pereira, 2005].

”Leur présence dans les données peut conduire à des estimateurs de paramètres biaisés et, suite à la réalisation de tests statistiques, à une interprétation des résultats erronée.” [Planchon, 2005].

Afin de détecter les valeurs aberrantes au niveau des variables numériques nous avons utilisé les boîtes à moustaches⁶ que nous avons obtenues par la fonction `boxplot()` de `seaborn` :

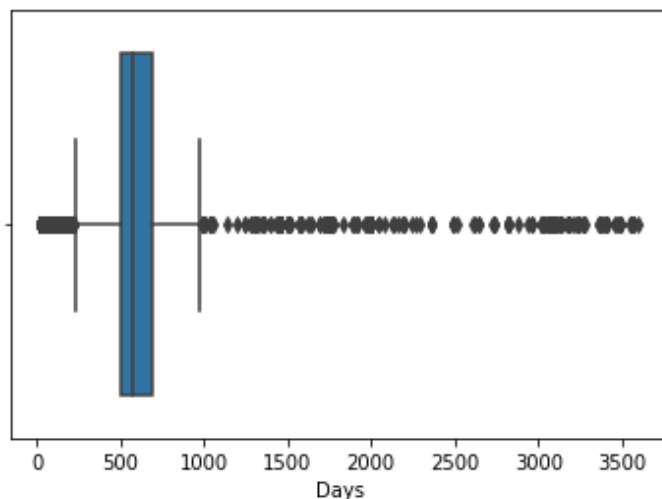


FIG. 42 : Boîte à moustaches sur la variables 'Days'

A travers la boîte à moustache, nous remarquons que la variable 'Days' prend des valeurs très proches de zéro, ces conteneurs ne sont pas considérés comme bloqués à cette période là, nous avons pu détecter 37 conteneurs dont le nombre de jours est inférieur à 22⁷ :

```
df['Days'].loc[df.Days < 22].count()
37
```

FIG. 43 : Capture d'écran du code python

Pour ce qui est des variables catégorielles nous avons examiné leurs modalités afin de détecter les valeurs aberrantes et les écarter dans but de maintenir les données les plus

⁶Représentation graphique permettant la détection des valeurs aberrantes dans un jeu de données.

⁷Seuil à travers lequel un conteneur rentre dans la liste.

pertinentes.

Test statistique de Hopkins :

Dans le but de tester l'aptitude d'une base de données à être partitionnée en plusieurs groupes, nous avons opté pour test statistique de Hopkins qui permet de voir si les données suivent une distribution uniforme.

Pour cela nous avons formulé le test d'hypothèse suivant :

$$\text{Hopkins statistics} \begin{cases} H > 0.5 \\ H < 0.5 \end{cases}$$

Si $H > 0.5$ cela veut dire que les données sont uniformément distribuées et donc la tendance au clustering est faible.

Si $H < 0.5$ nos données ont un caractère aléatoire spatial⁸ et ils ont tendance à être partitionnées en groupes.

Nous avons utilisé la fonction `hopkins()` prédéfini dans la bibliothèque `pyclustertend` pour appliquer le test de Hopkins au jeu de données dont nous disposons, nous avons obtenu les résultats suivants :

```
hopkins(df,df.shape[0])  
0.10782303247069311
```

FIG. 44 : Résultats du test de hopkins effectué sur python

H est largement inférieur à 0.5, l'hypothèse H_0 est donc rejetée, les données ne sont pas uniformément distribuées, elles ont donc tendance à être partitionnées en groupes.

4.2.3 Préparation des données

Afin d'obtenir des modèles d'apprentissage performants, il est inévitable de passer par l'étape de préparation des données car elle permet de canaliser nos données dans le but de maintenir les informations les plus pertinentes qui serviront de variables d'entrées aux modèles de clustering que nous développerons par la suite.

Traitement des valeurs aberrantes

Lors de cette étape, nous avons débuté par la suppression de l'ensemble des conteneurs correspondant au nombre de jours inférieur à 22.

La figure suivante représente les modalités de la variable "Standard Comments" que nous avons obtenu par la fonction `value_counts()` de python :

⁸Une structure spatiale totalement aléatoire est un ensemble de points répartis de manière totalement aléatoire dans un volume donné.

Force majeure	4340
*Legal actions started	327
*Notification Sent to Customer	316
*Customs documentation delays	265
*Cargo abandoned by customer, awaiting for stripping	145
*Container cannot be released due to missing OBL and/or payment from customer	60
NO COMMENTS	31
*Customs inspection	26
Called Customer	17
Sent message to origin	13
*Military Use	10
*Incorrectly updated in RKEM (state action plan in free text field)	6
*Container verified in good condition, will be used for export booking	5
*Notification to customer failed	4
Dsp: Lost/ Stolen	3
*Trucker/ Intermodal Vendor Contacted	1
Name: *Standard Comment, dtype: int64	

FIG. 45 : Modalités de la variable Standrad comments

Au niveau des commentaires nous avons constatés que certaines données peuvent être supprimées, c’est le cas des conteneurs retenus par un même client, ce dernier bloque 63% de l’ensemble des conteneurs dans la liste, ce client est traité d’une manière spéciale par le service client de Maersk, nous l’avons donc écarté de la base de données.

Les commentaires contenant ‘Intermodal vendor contacted’ ont été supprimés parce qu’actuellement Maersk Algérie ne dispose pas de plateformes logistiques et son activité se limite uniquement sur le transport maritime, il s’agit donc bien d’une erreur lors de la mise à jour des données sur système.

Nous avons aussi exclu les conteneurs qui ont été perdus ou volés de chez les clients.

Les commentaires suivants ‘ Notification Sent to Customer ‘ et ‘ Called Customer ‘ veulent dire que l’entreprise n’a toujours pas eu de retours de la part des clients, elle n’est donc pas informée de la raisons du blocage de ces conteneurs, nous avons écarté ces conteneurs afin de les traiter à part. Après les “Standard Comments” nous avons visualiser les modalités de la variable “Last Loc Name” qui est celle des ports de déchargement :

Algiers Terminal-Algeria	5029
Skikda Terminal-Algeria	180
Oran Terminal-Algeria	119
Entreprise Portuaire de Bejaia-Algeria	104
Annaba Terminal-Algeria	101
Algeciras - ML Terminal-Spain	36
Name: Last Loc Name, dtype: int64	

FIG. 46 : Modalités de la variable Last Loc Name

Nous avons détecté une anomalie, il s’agit de la présence du port d’Algésiras en Espagne, nous avons donc supprimer les conteneurs correspondant à cette valeur de la base de données.

Sélection des variables

Dans cette étape nous avons préféré opter pour la méthode delphi, cette dernière se base sur des avis d’experts pour effectuer des prévisions ou résoudre des problèmes.

La figure suivante illustre la démarche adoptée pour la sélection des variables à prendre

en considération en se basant sur la méthode delphi :

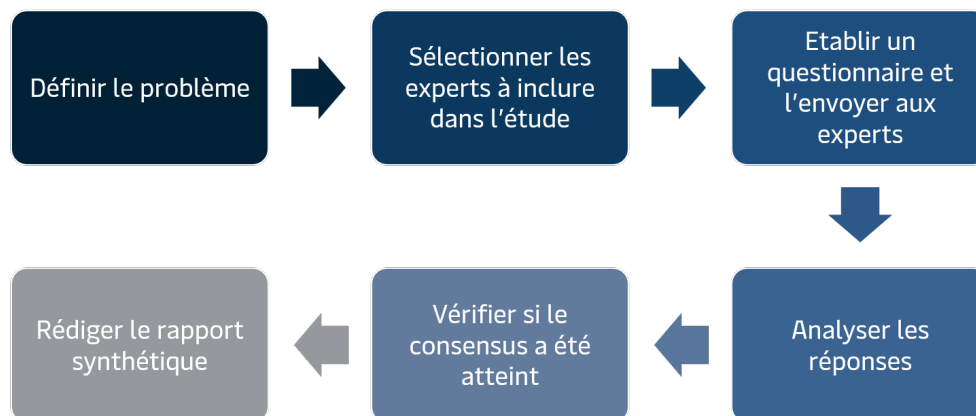


FIG. 47 : Méthode de sélection des variables

Nous présentons le déroulement de la démarche comme suit :

1. Définition du problème : Parmi les 17 variables présentes sur la base de données, quelles sont celles que nous retiendrons pour la construction du modèle de clustering ?
2. Sélection des experts : Il s'agit de l'ensemble des personnes concernées par la gestion des conteneurs bloqués :
 - Manager du service customer experience.
 - Responsable des agences présentes sur le territoire Algérien.
 - Chargé des réclamations.
 - Chargé des opérations d'export.
3. Entretiens : Nous avons effectué des entretiens en suivant la trame présentée au niveau des annexes. (Voir **Annexe G** p114)
4. Analyse des réponses : Après avoir effectué les entretiens avec l'ensemble des personnes concernées, nous avons regroupé les réponses comme suit :
 - Le nombre de jours est important car plus il est élevé, les pertes d'opportunités augmentent.
 - La variable indiquant le fait qu'un client a effectué l'échange de documentation avec la compagnie maritime doit être prise en considération car elle permet de voir si le client ne compte pas abandonner sa cargaison.
 - La taille et le type de conteneurs doivent être inclus dans le modèle de clustering car certains types de conteneurs comme les "Reefer" et les "Open" sont rares en Algérie.

- Le port de déchargement et la variable qui indique le dernier mouvement du conteneur sont aussi importants car ils permettent de localiser le conteneur et de confier sa gestion à l'agence concernée.
 - Les commentaires doivent être pris en considération car ils contiennent en général la raison derrière le blocage du conteneur.
 - Le type et le nombre de packages n'ont aucun impact sur la gestion des conteneurs.
 - La business unit de l'origine et le nom du navire ne sont pas liés au problème de rétention de conteneurs
 - Les variables (Consignee Customer, 1st TPDoc, Shipper Customer, UNNO) sont à écarter car ils représentent des numéros d'identification.
5. Vérification : Les avis des experts lors des entretiens ont clairement convergé, le consensus est donc vérifié, nous n'avons pas eu besoin d'établir un questionnaire pour départager.
6. Synthèse : Suite au déroulement de la démarche Delphi et en se basant sur des avis d'experts nous avons gardé les variables présentes sur la figure suivante :

```
Index(['Days', 'Eqp Size', 'Last Loc Code', 'Last Move', 'FRT RLS',  
      '*Standard Comment', 'Eqp Type'],  
      dtype='object')
```

FIG. 48 : Variables sélectionnées

Traitement des valeurs manquantes

Plusieurs options s'offrent à nous pour le traitement des données manquantes :

- Remplacer les valeurs manquantes par la moyenne.
- Supprimer les valeurs manquantes.
- Estimer les valeurs manquantes en effectuant une régression avec d'autres variables.
- Supprimer les variables contenant des valeurs manquantes.

Pour notre cas nous ne pouvons opter pour l'une des options précédentes car les données manquantes représentent le fait qu'un conteneur ne contient pas de produits dangereux ou spéciaux.

Il suffit donc de remplacer les valeurs manquantes par le chiffre 0 et effectuer un encodage pour le reste des valeurs afin de poursuivre l'analyse et fournir à la machine des valeurs numériques.

Encodage des variables catégorielle

Les modèles d'apprentissage automatique exigent que toutes les variables d'entrée et de

sortie soient numériques. Cela signifie que si les variables contiennent des données catégorielles, nous devons les encoder en chiffres avant de pouvoir ajuster et évaluer un modèle.

Les deux techniques les plus utilisés pour des projets de machine learning sont le “Label Encoding” et “One Hot Encoding” :

Label encoding : Cette approche est très simple et consiste à convertir chaque valeur d’une colonne en un nombre.

Exemple :

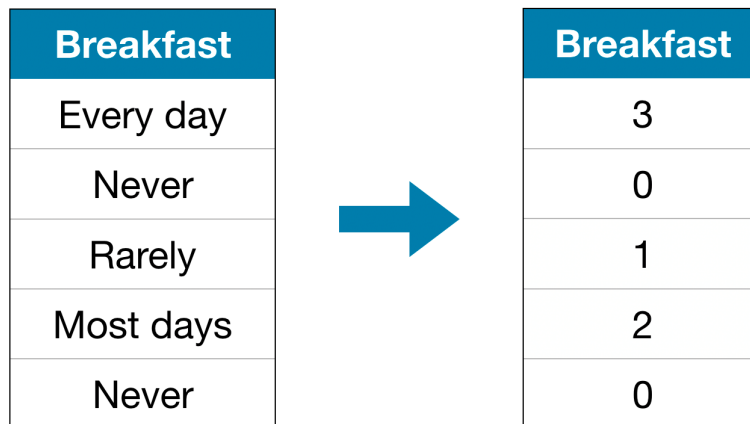


FIG. 49 : Exemple du label encoding

One Hot Encoding : Bien que Label Encoding soit simple, il présente l’inconvénient que les valeurs numériques peuvent être mal interprétées par les algorithmes de machine learning comme ayant une sorte d’hierarchie, ce problème est pris en considération par l’encodage one hot qui transforme chaque modalité en une variable contenant des valeurs binaires 1 ou 0.

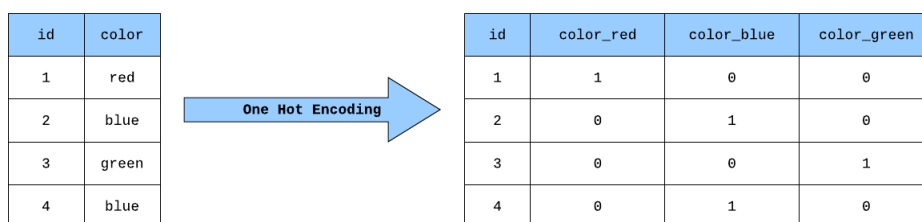


FIG. 50 : Exemple du one hot encoding

Pour le cas des variables catégorielles de notre base de données nous avons préféré opter pour le “Label Encoding” en raison de la présence de nombreuses variables catégorielles qui contiennent plusieurs modalités , en plus il s’agit d’un projet de clustering et donc il n’y a pas de risque pour que les algorithmes comprennent mal les valeurs introduites à l’entrée.

4.2.4 Modélisation des données

Après la fin des étapes de préparation des données, nous passons maintenant à la modélisation, nous allons présenter l'ensemble des algorithmes de clustering que nous avons utilisés pour le projet.

Nous allons donc tester l'ensemble des modèles définis au niveau du chapitre précédent afin de pouvoir sélectionner celui qui donne le meilleur résultat.

Application du K-Means :

L'initialisation de l'algorithme KMeans nécessite la précision du nombre de clusters à la sortie, pour cela nous avons utilisés deux méthodes afin de le déterminer :

La première consiste à varier le nombre de clusters et visualiser le coefficient de silhouette, nous avons obtenu les résultats suivants :

```

For n_clusters=2, The Silhouette Coefficient is 0.8289589085461424
For n_clusters=3, The Silhouette Coefficient is 0.8198065360797498
For n_clusters=4, The Silhouette Coefficient is 0.734869934753047
For n_clusters=5, The Silhouette Coefficient is 0.7300522591200911
For n_clusters=6, The Silhouette Coefficient is 0.69698822274648
For n_clusters=7, The Silhouette Coefficient is 0.6908566891022916
For n_clusters=8, The Silhouette Coefficient is 0.6863259320183215
For n_clusters=9, The Silhouette Coefficient is 0.7005541222181948
For n_clusters=10, The Silhouette Coefficient is 0.6972308469852013
For n_clusters=11, The Silhouette Coefficient is 0.7059673987516345
For n_clusters=12, The Silhouette Coefficient is 0.7099604532304102
For n_clusters=13, The Silhouette Coefficient is 0.6774049772268775
For n_clusters=14, The Silhouette Coefficient is 0.682523341921712
    
```

FIG. 51 : Résultat obtenus suite à la variation du coefficient de silhouette

Par la suite nous avons confirmé ce résultat par la méthode du coude, en variant l'inertie en fonction du nombre de clusters, cette méthode est illustrée par le graphique suivant :

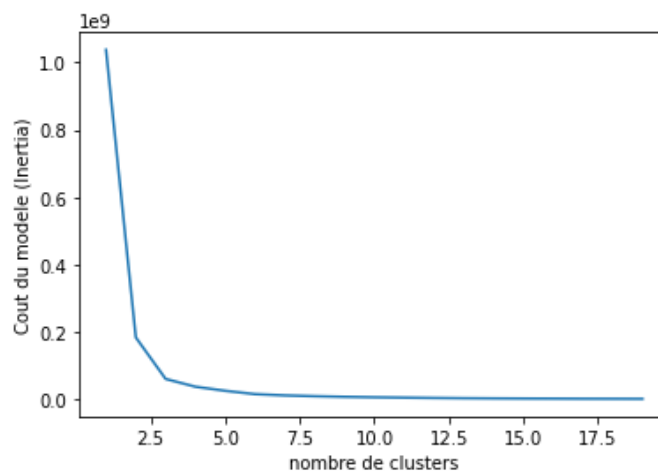


FIG. 52 : Graphique de l'inertie en fonction du nombre de clusters

Nous pouvons voir à travers les deux graphiques précédents que le nombre de cluster optimal est de 3 car la cassure du coude se trouve correspond à $k = 3$ et le coefficient de

silhouette est égal à 0.81 pour cette même valeur.

Après la détermination du nombre de clusters, nous avons initialiser l'objet KMeans à partir de la bibliothèque Sklearn tout en précisant le nombre de cluster à 3, par la suite nous avons entraîné le modèle en introduisant la base de données. La dernière étape de cette application consiste à utiliser la fonction `predict()` afin de retourner le numéro de cluster sous forme de tableau numpy.

Application du minibatch :

Comme Minibatch n'est qu'une variante de l'algorithme K-Means, il suffit donc d'introduire le nombre de clusters à ce dernier afin que le modèle puisse être entraîné, pour cela nous avons opter pour la méthode de variation du coefficient de silhouette qui a donné un nombre de cluster égal à 3 avec une valeur de 0,82 pour le coefficient de silhouette.

Application du DBSCAN :

L'entraînement de DBSCAN est facilité par Scikit-Learn tout comme les autres algorithmes de clustering, pour cela il faudrait introduire deux paramètres d'entrée parmi eux "eps" qui représente la distance entre deux points du jeu de donnée, celle-ci sera prise en considération par l'algorithme DBSCAN dans la construction des groupes de donnée. Afin de choisir le bon epsilon nous avons utilisé la distance euclidienne : soient $p = (p_1, \dots, p_n)$ et $q = (q_1, \dots, q_n)$:

$$d(\mathbf{p}, \mathbf{q}) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_i - q_i)^2 + \dots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}.$$

Afin d'optimiser eps, nous avons calculer pour chaque point du jeu de données la distance avec son voisin le plus proche, par la suite on fixe un eps tel que :

- Une grande partie des points ait une distance inférieure à eps.
- Une grande partie des points ait au moins un voisin dans son eps-voisinage.

La bibliothèque Scikit-Learn met à disposition les méthodes de la classe NearestNeighbors qui permettent de déterminer les voisins les plus proches ainsi que les distances.

Application de la classification ascendante hiérarchique :

Pour appliquer la classification ascendante hiérarchique, nous avons suivi la même méthode que pour le KMeans afin de choisir le nombre de clusters à introduire. Nous avons obtenu le résultat suivant :

```
For n_clusters=2, The Silhouette Coefficient is 0.8275453293863171
For n_clusters=3, The Silhouette Coefficient is 0.8153282083346876
For n_clusters=4, The Silhouette Coefficient is 0.7416726648037092
For n_clusters=5, The Silhouette Coefficient is 0.678461540551018
For n_clusters=6, The Silhouette Coefficient is 0.6779185102309635
For n_clusters=7, The Silhouette Coefficient is 0.6738962655534098
For n_clusters=8, The Silhouette Coefficient is 0.6738872836062976
For n_clusters=9, The Silhouette Coefficient is 0.6770187821305926
For n_clusters=10, The Silhouette Coefficient is 0.6401952766964123
For n_clusters=11, The Silhouette Coefficient is 0.6400300924290816
For n_clusters=12, The Silhouette Coefficient is 0.639750502199658
For n_clusters=13, The Silhouette Coefficient is 0.6384886978549448
For n_clusters=14, The Silhouette Coefficient is 0.6424770020928194
```

FIG. 53 : Résultat obtenus suite à la variation du coefficient de silhouette en utilisant la CAH

Nous avons préféré fixer le nombre de clusters à 3 pour plus de précision car la différence est de 0.01 seulement entre les deux valeurs.

Par la suite nous avons instancié l'objet `AgglomerativeClustering()` de la bibliothèque Scikit-Learn en précisant le nombre de clusters déjà déterminé ainsi que le type de distance à utiliser (euclidienne).

Après l'entraînement du modèle CAH ; nous avons visualiser le dendrogramme à l'aide de la fonction `dendrogram` du module Scipy, nous avons obtenu le résultat suivant :

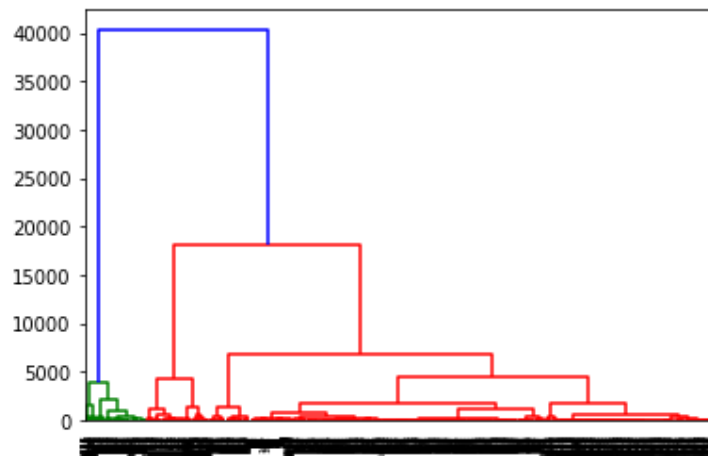


FIG. 54 : Visualisation du dendrogramme

MeanShift

Pour entraîner cet algorithme, nous avons utilisé la fonction `estimate_bandwidth()` qui a permis d'optimiser les paramètres d'entrées de l'algorithme présent. Le modèle donne 4 différentes classes de conteneurs.

4.2.5 Evaluation des modèles

Afin d'évaluer la performance des modèles que nous avons obtenu en utilisant les 5 algorithmes (K-moyennes, Minibatch, CAH, DBSCAN et Meanshift), nous avons utilisé les trois métriques que nous avons expliqué au niveau du Chapitre 3, ces derniers sont :

- Coefficient de silhouette.
- Indice de Davies-Bouldin.
- Indice de Calinski Harabasz.

Ces indices sont déjà implémentés au niveau du module python sklearn.metrics, nous les avons donc appliquer en instanciant les objets : silhouette_score, calinski_harabasz_score, davies_bouldin_score.

Pour le coefficient de silhouette nous avons obtenu le résultat suivant :

Modèles	Score	Classement
K-Means	0.820	1
DBSCAN	0.588	5
MiniBatch	0.817	2
CAH	0.815	3
Meanshift	0.797	4

TAB. 6 : Tableau des résultats obtenus suite à l'évaluation par le coefficient de silhouette.

Pour l'indice de Calinski-Harabasz :

Modèles	Score	Classement
K-Means	13428.137	1
DBSCAN	879.518	5
MiniBatch	13419.196	2
CAH	13259.261	3
Meanshift	10369.925	4

TAB. 7 : Tableau des résultats obtenus suite à l'évaluation par l'indice de Calinski-Harabasz.

Pour l'indice de Davies-Bouldin : Contrairement aux deux indices précédents, la valeur la plus basse correspond au modèle le plus performant.

Modèles	Score	Classement
K-Means	0.320	3
DBSCAN	1.052	5
MiniBatch	0.315	2
CAH	0.313	1
Meanshift	0.469	4

TAB. 8 : Tableau des résultats obtenus suite à l'évaluation par l'indice de Davies-Bouldin.

Afin de choisir le modèle le plus performant tout en tenant compte de ces trois indices, nous avons affecté un poids pour chaque classement tel que le poids du classement 1 est de 5 et celui du classement 5 est de 1.

Modèles	classements	Occurences	Score
K-Means	1	2	13
	2	0	
	3	1	
	4	0	
	5	0	
DBSCAN	1	0	3
	2	0	
	3	0	
	4	0	
	5	3	
Minibatch	1	0	12
	2	3	
	3	0	
	4	0	
	5	0	
CAH	1	1	7
	2	0	
	3	2	
	4	0	
	5	0	
Meanshift	1	0	6
	2	0	
	3	0	
	4	3	
	5	0	

TAB. 9 : Tableau des résultats d'évaluation.

L'algorithme K-means possède le score le plus élevé (13 points) suivi du MiniBatch avec 12 points, nous pouvons visualiser les scores à travers le diagramme en bâton suivant :

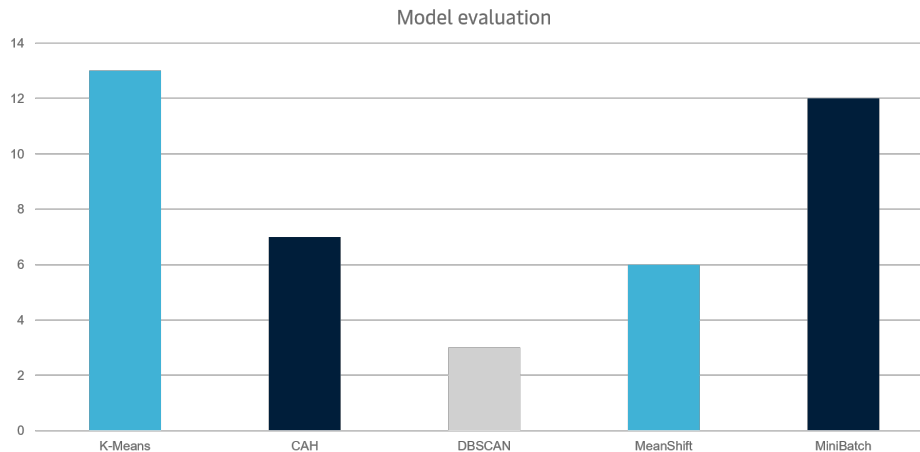


FIG. 55 : Visualisation des résultats d'évaluation par le diagramme en bâton.

K-means donne des résultats très convaincants sur les trois indices, nous l'avons donc maintenu afin de l'utiliser dans le déploiement de l'outil d'aide à la décision.

Nous avons pu visualiser les différents clusters obtenus en utilisant un scatter-matrix⁹ généré à l'aide de la fonction `pairplot()` de `seaborn`. (Voir **Annexe H** p115)

4.3 Segmentation des clients

Afin de segmenter les différents clients présents dans de la base de données des conteneurs en souffrance, nous avons opté pour la même démarche que celle des conteneurs, pour cela nous avons pu extraire une base de données clients à partir de la base de données brute, dans le but de répartir les individus sur plusieurs groupes appelés clusters.

Nous allons donc dérouler la démarche CRISP-DM :

4.3.1 Compréhension du cas

Après avoir effectué la segmentation des conteneurs, nous allons s'intéresser à celle des clients et cela dans le but de comprendre les caractéristique de chaque type de clients selon plusieurs critères, parmi eux : le nombre de conteneurs en souffrance généré par le client, le nombre de ports où se situent ces conteneurs, le nombre de connaissances etc...

Cette segmentation permettra de caractériser les clients présents dans la base de données avec plus de précision.

⁹ Représentation de données dépendant de plusieurs variables.

4.3.2 Compréhension et préparation des données

Contrairement à celle des conteneurs, la base de données des clients est composée majoritairement de variables numérique et une seule variable catégorielle qui représente le numéro d'identification des clients, nous avons pu voir certaines priorités de la base de données en utilisant la fonction `info()` de Pandas :

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 287 entries, 0 to 286
Data columns (total 7 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   Row Labels                                286 non-null    object
1   Max of Days                               287 non-null    int64
2   Min of Days                               287 non-null    int64
3   Distinct Count of Last Loc Name          287 non-null    int64
4   Count of Equipment Number               287 non-null    int64
5   Distinct Count of 1st TPDoc             287 non-null    int64
6   Distinct Count of Shipper Customer       287 non-null    int64
dtypes: int64(6), object(1)
memory usage: 15.8+ KB
```

FIG. 56 : Visualisation des informations sur la base de données clients

Row labels : Il s'agit du numéro d'identification, il est unique pour chaque client.

Max of Days : C'est le nombre de jours correspondant à la durée de rétention du conteneur le plus ancien.

Min of Days : C'est le nombre de jours correspondant à la durée de rétention du conteneur le moins ancien.

Distinct Count of Last Loc Name : Le nombre de ports au niveau desquels le client retient des conteneurs.

Count of Equipment Number : Il s'agit du nombre de conteneurs retenu par le même client.

Distinct Count of 1st TPDoc : Le nombre de connaissances à travers lesquels un client à générer des conteneurs.

Distinct Count of Shipper Customer : Le nombre de fournisseurs avec lesquels le client a généré des conteneurs en souffrance.

Valeurs manquantes : Afin de vérifier si notre jeu de donnée contient des variables manquantes nous avons utilisé la méthode `isnull()` et la fonction `sum()` qui ont permis de comptabiliser les valeurs manquantes pour chaque variable, nous avons obtenu les résultats suivants :

La base de données contient donc une seule valeur manquante qui correspond au numéro d'identification du client, il suffit de générer un code client pour la compléter.

```
Row Labels      1
Max of Days    0
Min of Days    0
Distinct Count of Last Loc Name  0
Count of Equipment Number      0
Distinct Count of 1st TPDoc    0
Distinct Count of Shipper Customer  0
dtype: int64
```

FIG. 57 : Visualisation des valeurs manquantes sur python

Valeurs aberrantes :

Afin de détecter de potentiels valeurs aberrantes ou erreur nous avons visualiser la distribution de la variable qui correspond au nombre de conteneurs par la fonction `boxplot()` de `seaborn`, le résultat est illustré sur la figure suivante :

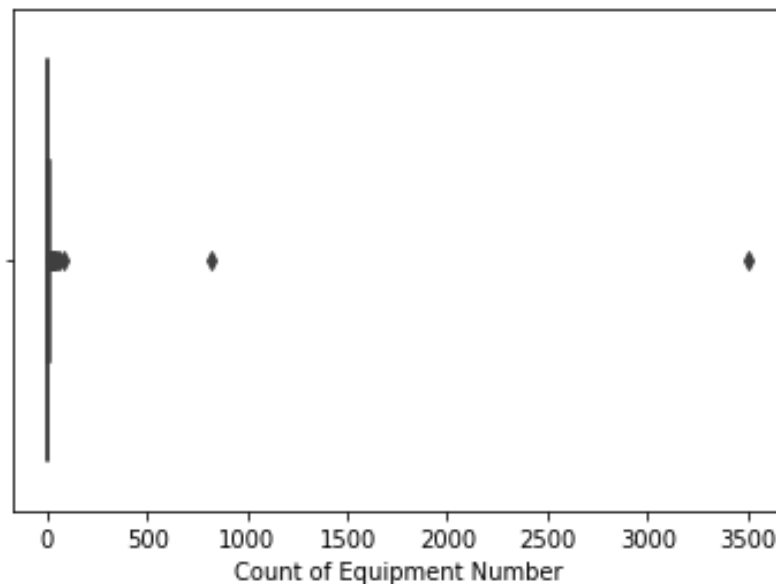


FIG. 58 : Boite à moustaches sur la variable Count of Equipment Number

Nous remarquons la présence d'un client qui retient près de 3500 conteneurs, il s'agit donc d'un client spéciale qui n'est pas géré de la même manière que les autres clients, pour cette raison nous avons préféré l'écarter de la base de données, car il il risque de biaiser les résultats des algorithmes de partitionnement.

Sélection des variables :

Dans le but de sélectionner les variables les plus pertinentes à l'entraînement du modèle de clustering, nous avons avons afficher les variances des variables à l'aide de la fonction `var()` :

```
df.var()
Max of Days          1.177819e+06
Min of Days          1.172686e+06
Distinct Count of Last Loc Name  2.503532e-02
Count of Equipment Number  4.665947e+04
Distinct Count of 1st TPDoc  4.316062e+01
Distinct Count of Shipper Customer  3.937634e-01
dtype: float64
```

FIG. 59 : Capture d'écran python des variances pour les variables clients

Nous remarquons que les variables relatives au nombre de ports et au nombre de fournisseurs possèdent une variance presque nulle.

Les deux variables Max of Days et Min of Days ont la même variance, nous avons donc écarté ces trois variables dans le but d'obtenir un modèle de clustering plus performant.

4.3.3 Modélisation des données

Au niveau de cette étape nous avons entraîné les cinq algorithmes déjà utilisés dans la première section.

Pour les algorithmes K-Means, Minibatch et CAH nous nous sommes basés sur le coefficient de silhouette et la méthode des coudes afin de fixer le nombre de clusters optimal à introduire.

Modèles	Nombre de clusters introduit	Nombre de clusters obtenus
K-Means	3	/
DBSCAN	/	1
CAH	3	/
MeanShift	/	6
Minibatch	3	/

TAB. 10 : Tableau contenant le nombre de clusters introduit ou obtenu lors de la phase de modélisation.

4.3.4 Evaluation des modèles

Dans le but de choisir le modèle que nous allons maintenir pour la suite de notre travail, nous avons opté pour la même démarche utilisée au niveau la section 1.

Nous avons maintenu les indices d'évaluation suivants :

- Coefficient de silhouette.
- Indice de Calinski-Harabasz.
- Indice de Davies Bouldin.

Nous avons effectué l'évaluation à l'aide des fonctions python présentées au niveau de la figure suivante, ces dernières retournent la valeur de l'indice relative à chaque modèle :

```
#Silhouette
print("Silhouette Coefficient KMeans: %0.3f" % metrics.silhouette_score(df, kmeans.labels_))
print("Silhouette Coefficient Agglomerative: %0.3f" % metrics.silhouette_score(df, cah.labels_))
print("Silhouette Coefficient MiniBatch: %0.3f" % metrics.silhouette_score(df, minibatch.labels_))
print("Silhouette Coefficient DBSCAN: %0.3f" % metrics.silhouette_score(df, db.labels_))
print("Silhouette Coefficient MeansShift: %0.3f" % metrics.silhouette_score(df, ms.labels_))
```

FIG. 60 : Fonctions python utilisés pour le calcul des indices d'évaluation

Le tableau suivant résume les résultats de l'évaluation des 5 modèles entraînés :

Modèles	Coefficient de silhouette	Indice de Calinski-Harabasz	Indice de Davies Bouldin
K-Means	0.69	1094	0.42
CAH	0.68	1036	0.42
Minibatch	0.69	855	0.48
DBSCAN	/	/	/
MeanShift	0.46	725	0.45

TAB. 11 : Tableau montrant les résultats d'évaluation du modèle de clustering pour les clients

Nous pouvons facilement remarquer que le modèle entraîné par K-Means donne des résultats satisfaisants sur les trois indicateurs en comparant avec les 4 autres modèles. Comme pour le cas de la segmentation des conteneurs nous avons opté pour K-Means pour la phase de déploiement.

La figure suivante représente la visualisation des différents clusters donné par K-Means :

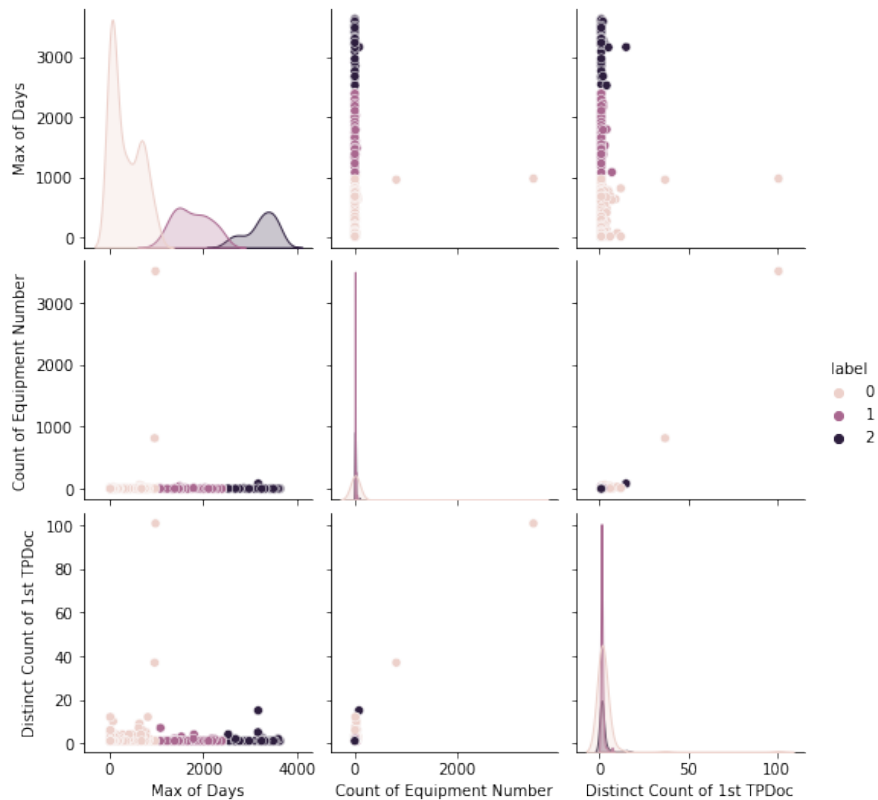


FIG. 61 : Scatter-matrix après avoir effectué le clustering par K-means

4.4 Déploiement

Dans cette étape, nous allons expliquer l'intérêt derrière les deux segmentations que nous avons effectué par les méthodes d'apprentissage non supervisé dans le but de construire un outil d'aide à la décision qui permettra d'optimiser le traitement des différents cas de conteneurs en rétention.

4.4.1 Classification ABC

Avant de présenter les résultats du clustering, nous avons traité le cas des conteneurs que nous avons écarté de la base de données lors de la segmentation des conteneurs, ces derniers représentent les conteneurs retenus par des clients qui n'ont toujours pas fourni à Maersk la raison derrière la rétention des conteneurs, nous avons pu identifier ces conteneurs à l'aide de la variable "Standard Comments" présentes sur la base de données brutes.

Après avoir filter la base de données nous avons retenu les observations qui correspondent aux deux modalités suivantes :

```
df['*Standard Comment'].value_counts()

*Notification Sent to Customer    316
Called Customer                    17
Name: *Standard Comment, dtype: int64
```

FIG. 62 : Capture d'écran python des modalités de la variable Standard Comments

Les clients qui retiennent 94% des conteneurs ont été notifié par l'envoi d'un e-mail suivant la SOP expliqué dans le chapitre 2, les 6% restants ont été appelés.

La base de données contient donc 333 conteneurs en rétention, pour pouvoir les gérer nous les avons classés par le recours à la méthode ABC qui va nous permettre de situer les différents cas présents par rapport au nombre de jours de rétention.

Nous avons effectué la classification par le langage python en injectant la première base de données pré-traitée au niveau de la segmentation des conteneurs, il suffit donc de la filtrer afin de pouvoir l'utiliser pour classifier ces conteneurs comme le montre le schéma suivant :

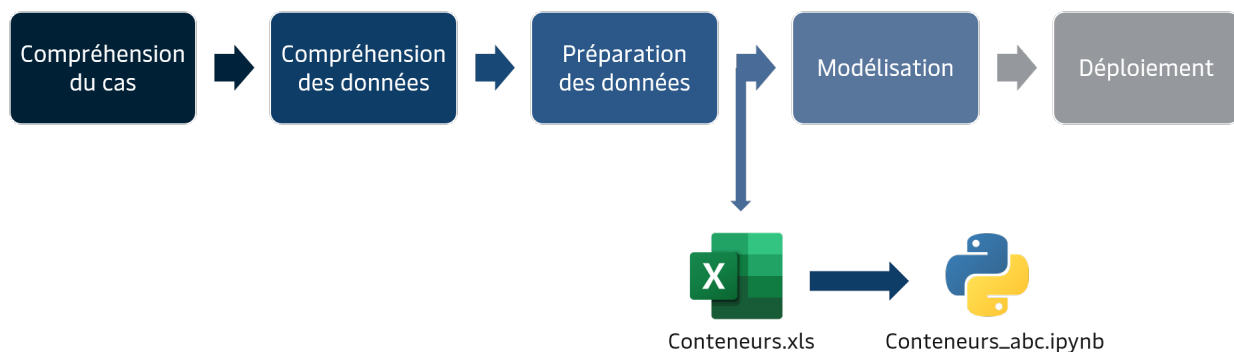


FIG. 63 : Schéma expliquant la provenance des données d'entrées pour la classification ABC

Au niveau du code python nous avons utilisé les bibliothèque pandas et numpy expliquée dans le chapitre 3, la figure suivante représente le code python que nous avons implémenté :

```
df['cumul'] = df['Days'].cumsum()
df['%cumulStocks'] = round(100 * df['cumul'] / df['Days'].sum(), 3)
classes = ['A', 'B', 'C']
df['classe'] = pd.cut(df['%cumulStocks'], 3, labels=classes)
df.to_excel('ABCClass.xlsx')
```

FIG. 64 : Capture d'écran python montrant le code implémenté pour la classification ABC

La fonction `cumsum()` de numpy retourne la somme cumulée des éléments tout le long de l'axe donné, par la suite nous avons calculé le pourcentage cumulé, en dernier lieu et en utilisant la fonction `cut` de pandas nous avons tiré les trois classes A B C :

Classe A : Durée de rétention entre 179 et 688 jours, représente 7,22% du total, ces conteneurs risquent d'être abandonné par le client concerné et nécessite un suivi rigoureux pour essayer d'avoir les raisons pour lesquels ces conteneurs sont bloqués.

Classe B : Cette classe représente 37,86% du total, elle est caractérisé par un nombre de jours compris entre 28 et 179, comme pour la classe A cette catégorie nécessite un suivi régulier pour comprendre les cas en questions et pouvoir les traiter.

Classe C : Elle représente 54,92% de l'ensemble des conteneurs présents dans la liste, la durée de rétention est comprise entre 22 et 28 jours, cette catégorie ne demande pas un suivi régulier pour le moment car une grande partie des conteneurs est restituée après le 28ème jour en général.

4.4.2 Résultats du clustering

Suite à la validation du modèle de partitionnement des conteneurs, nous avons obtenu trois groupes distinct de conteneurs, chaque groupe possède des caractéristiques différentes, le diagramme suivant représente la répartition des conteneurs sur ces trois groupes :

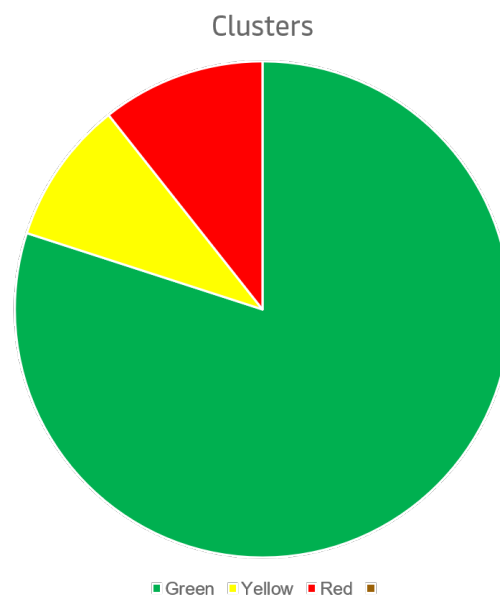


FIG. 65 : Diagramme en camembert représentant la répartition des conteneurs sur les clusters

Nous résumons à travers le tableau suivant les caractéristiques de chaque groupe de conteneurs :

Clusters	Critères	Valeurs
Green	Nombres de jours	44-1090
	Commentaires (Causes)	Retards liés à la documentation
	Type/size	1286 DRY/ 26 FLAT/ 17 OPEN/ 5 REEFER
Yellow	Nombres de jours	2522 - 3635
	Commentaires (Causes)	Cas légal
	Type/size	152 DRY / 1 FLAT / 1 OPEN /2 REEFER
Red	Nombres de jours	1180 - 2402
	Commentaires (Causes)	Conteneurs abandonnés par les clients
	Type/size	166 DRY / 1 FLAT /11 REEFER

TAB. 12 : Tableau récapitulatif du résultat de clustering des conteneurs.

Green containers : C'est la classe qui contient le plus grand nombre de conteneurs (79% des conteneurs en rétention), les conteneurs appartenant à cette classe possède un nombre de jours de rétention assez bas, souvent lié à des problèmes de documentation entre les clients et la douane, cette classe de conteneurs est caractérisée aussi par un nombre important de conteneurs Flat et Open qui sont très rare en Algérie et sont nécessaires pour les opérations d'exportation.

Yellow containers : Cette classe de conteneurs possède un nombre de jours de rétention très important, il s'agit des cas "Légal" cela signifie que Maersk a déjà entamé une procédure judiciaire dans le but de récupérer ses conteneurs.

Red containers : La majorité de ces conteneurs présents au niveau de cette catégorie ont été abandonné par le client au niveau du port pour une des raisons suivantes :

- Problèmes de documentation bancaire.
- Clients non joignables.
- Les clients ne possèdent pas les documents ou les moyens financiers nécessaires pour effectuer l'échange avec la compagnie maritime.

Actuellement, il y a une pénurie de conteneurs Reefer en Algérie et particulièrement chez Maersk, cette catégorie contient un nombre important de conteneurs réfrigérés (61% des conteneurs reefer en rétention). Pour le cas des clients nous avons également retenu le modèle entraîné par l'algorithme K-Means, nous présentons à travers le tableau suivant les résultats relatifs à cette segmentation :

Clusters	Critères	Valeurs
Red	Nombres de jours	22 - 1388
	Nombre moyen de conteneurs	27
	Nombre maximal distinct de connaissements	101
Green	Nombres de jours	597 - 2402
	Nombre moyen de conteneurs	4
	Nombre maximal distinct de connaissements	4
Yellow	Nombres de jours	2522 - 3635
	Nombre moyen de conteneurs	3
	Nombre maximal distinct de connaissements	15

TAB. 13 : Tableau récapitulatif du résultat de clustering des clients

Red customer : Cette classe de clients est caractérisée par un nombre élevé de conteneurs en rétention, il est de même pour le nombre de connaissance cela veut dire que les clients de cette classes ont tendance à générer des conteneurs en souffrance à chaque fois opération d'import effectué avec Maersk.

Green customer : Les clients appartenant à cette catégorie retiennent en moyenne 4 conteneurs sur un nombre maximal de 4 connaissements c'est donc des clients qui ne génère pas de conteneurs immobiles de manière répétitif, il s'agit en général des entreprises de taille moyenne.

Yellow customer : En général cette classe contient des clients déjà blacklistés, ils sont caractérisés pas un nombre de conteneurs et un nombre de jours de rétention très élevé. Afin de mieux catégoriser les clients et les conteneurs nous avons utilisé l'estimation faite au niveau du Chapitre 2 pour pouvoir estimer les perte qu'engendre chaque segment de clients et de conteneurs, les résultats sont représenté sur la tableau suivant :

Résultats	Clusters	Nombres de conteneurs	Pertes estimées	Pertes/Conteneurs
Conteneurs	Green	1333	34 576 835,67 €	25 939
	Yellow	156	19 750 498,93 €	126 605
	Red	178	11 573 217,91 €	65018

TAB. 14 : Tableau montrant l'estimation des pertes pour chaque cluster de conteneurs

Nous remarquons à travers ce tableau que les conteneurs faisant partie du cluster Yellow causent plus de pertes comparant aux deux autres clusters cela est du au fait que les cas "Legal" prennent beaucoup de temps, les conteneurs appartenant au cluster "Green" sont les moins coûteux avec un ratio pertes/conteneurs de 25 939.

4.4.3 Mode opératoire

Dans cette section nous expliquons les étapes d'utilisation de l'outil que nous avons développé ainsi que nos recommandations par rapport aux traitements des différents cas de conteneurs en souffrance.

Le schéma suivant représente les différentes étapes à travers lesquelles passent les données que nous avons utilisés :

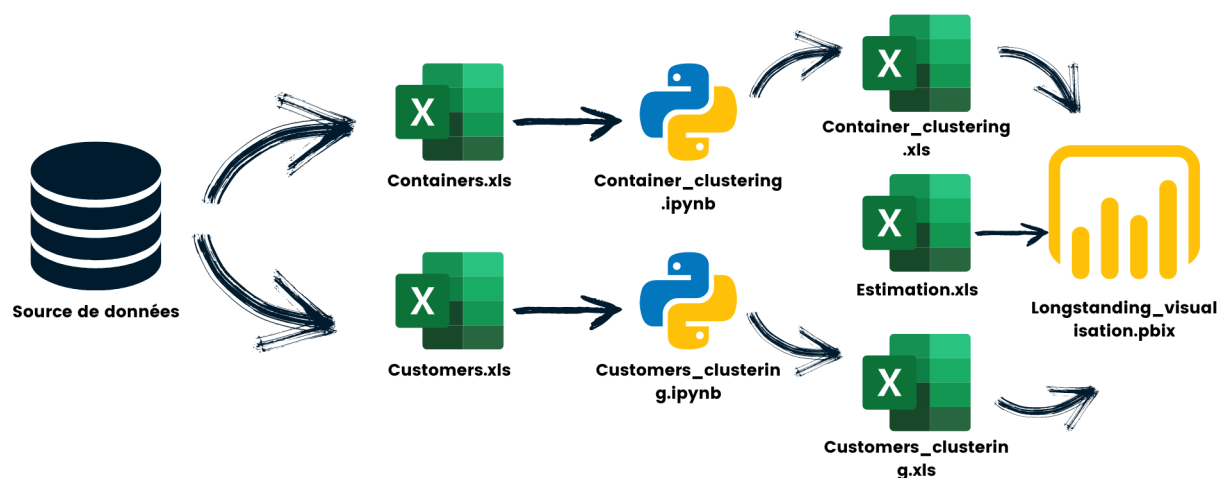


FIG. 66 : Schéma récapitulatif de la solution

En premier lieu, l'utilisateur tire un rapport excel du système de tracking de conteneurs en sélectionnant les variables déjà présentées au niveau de la première section, par la suite l'utilisateur extrait une base de données contenant les caractéristiques des clients comme (Nombre de conteneurs, nombre de connaissances, etc...), ces deux fichier excel serviront de données d'entrées au script python implémenté sur Google Colaboratory, après l'exécution du code python deux fichiers excel seront extraits directement, ils contiennent la variable précisant les clusters associés, ces deux même fichiers accompagnés de celui de l'estimation des pertes seront injecter sur Power BI dans le but de visualiser l'état des conteneurs en souffrance à l'instant t. Nous pouvons résumer le

mode opératoire en visualisant le processus modélisé par BPMN à travers la figure suivante :

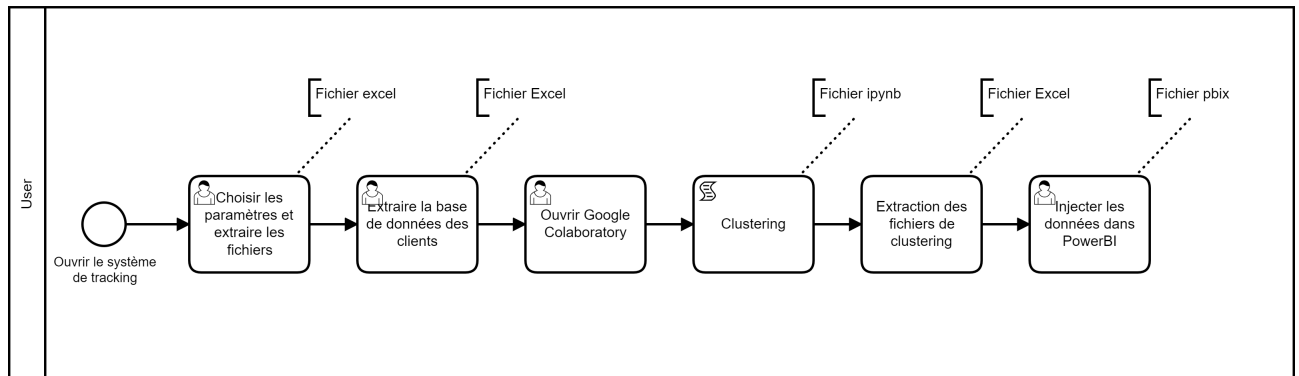


FIG. 67 : Modélisation BPMN du processus mis en place

4.5 Mise en place du tableau de bord :

A travers la démarche choisie, nous déterminerons les indicateurs de performance les plus pertinents, en corrélation avec les objectifs du département Customer Service de Maersk Algérie. Elle servira donc de ligne directrice tout au long de la conception et de la construction du TDB sur lequel seront présentés et suivis ces indicateurs.

4.5.1 Choix de la démarche de construction du tableau de bord

La littérature propose des différentes approches et méthodes de construction de tableaux de bord que nous avons déjà présenté dans la partie théorique.

Notre choix s'est porté sur la méthode GIMSI car tout d'abord celle-ci est plus adaptée aux structures aplaties favorisant l'initiative et les échanges latéraux. Contrairement aux deux autres méthodes qui appliquent une hiérarchie (top-down) dans leurs processus de construction.

Aussi, c'est une démarche coopérative, misant sur une incorporation douce à la complexité ambiante, elle est fondée sur un principe simple de responsabilisation, d'autonomie et de communication de l'ensemble des acteurs, décideurs à part entière.

Offrant la possibilité d'utiliser au mieux les outils de communication et de partage proposés par l'entreprise pour la mise en place de l'intelligence collective, le choix de la démarche GIMSI s'adapte parfaitement aux perspectives de l'équipe CX pour la coordination de leurs objectifs et comportements.

4.5.2 Déroulement de la méthode GIMSI :

Lors de la partie suivante, nous entamerons la seconde phase de la démarche GIMSI, la phase de "conception", puisque la première phase, la phase "d'identification" qui s'appuie sur l'analyse de l'environnement économique de l'entreprise (étape 1), et

l'identification de la structure organisationnelle (étape 2) à été déjà abordée lors du chapitre 1 et 2.

La phase de conception

Étape 3: Définition des objectifs

Ceci revient à définir l'ensembles des objectifs du processus de gestion des conteneurs immobiliers. Pour ce faire, nous avons consulté des documents formels de l'entreprise : rapports annuels de restitution des conteneurs immobiliers, plans directeurs..Etc. D'un autre coté, nous avons assisté à quelques rencontres des dirigeants, organisé des ateliers de brainstorming avec quelques membres de l'équipe CX pour en tirer indicateurs mais aussi les préoccupations d'amélioration et des objectifs.

Ainsi, en se basant sur l'analyse déjà effectuée dans le chapitre 2, qui décline les priorités globales du département, nous avons déduit les objectifs suivant :

1. **Objectif(1)** : Avoir un gain en temps en améliorant la réactivité de l'équipe CX.
2. **Objectif(2)** : Recenser le nombre de conteneurs immobiliers.
3. **Objectif(3)** : Estimer les pertes liés a la retenue des conteneurs.
4. **Objectif(4)** : Amélioration des restitutions annuelles d'au moins 30%.

Étape 4 : Construction du tableau de bord

Une fois l'étape précédente achevée, la définition de la structure du TDB est établie à fin de s'assurer qu'il ne se résume pas en un assemblage aléatoire d'indicateurs, mais à un outil d'aide à la décision garantissant un agencement précis entre les indicateurs choisis et les objectifs cibles de l'entreprise

Par conséquent, la structure du tableau de bord dépend de la structure de ses indicateurs qui ont pour rôle de mesurer la situation actuelle et déclencher une réaction du décideur en fonction de l'objectif cible. Précisons que la réaction peut être de ne rien faire, mais dans tout les cas, ce dernier déclenchera une prise de décision.

Les KPIs en question peuvent être :

- Des indicateurs d'alerte, qui signalent un dysfonctionnement impliquant des interventions.
- Des indicateurs d'équilibre, qui informent l'avancement par rapport aux objectifs pouvant induire des actions correctives.
- Des indicateurs d'anticipation, qui permettent de reconsidérer la stratégie préalablement choisie en exerçant des actions préventives.

Etape 5 : Choix des indicateurs de performance

Cette partie va englober la sélection d'indicateurs de mesures pertinents du processus en question,l'analyse menant à la sélection et construction de ces derniers s'avère être une

étape critique. Cependant, les KPIs doivent être soigneusement choisis et soumis tout au long de cette analyse aux critères définis par la démarche GIMSI :

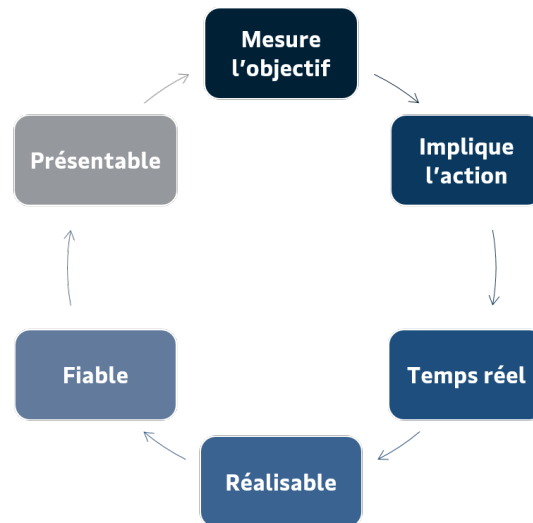


FIG. 68 : Les critères d'un bon indicateur définis par la démarche GIMSI

- **Mesure l'objectif** : Il mesure la performance selon un ou plusieurs objectifs préalablement définis dans l'étape précédent.
- **Implique l'action** : un indicateur a pour objectif d'inciter quelque part l'utilisateur à prendre des décisions et réagir face à l'information portée.
- **Temps réel** Le TDB ne présentera que des indicateurs qui sont adaptés aux changements instantanés favorisant la prise de décision.
- **Réalisable** : Il va falloir construire le KPI avec les données disponibles dans le système d'information de l'entreprise.
- **Fiable** : Un décideur n'utilise jamais un indicateur que dans la mesure où il le juge fiable, il doit avoir donc accorder une confiance totale à son TDB.
- **Présentable** : Le KPI doit être facilement représenté, de manière à ce que le sens porté soit compris le mieux possible.

De manière plus spécifique à notre problématique et selon les accès que nous avons eu à notre disposition, nous avons choisi les indicateurs résumés ci-dessous :

- Le nombre de conteneurs en souffrance.
- Le nombre de clients engendrant la rétention des conteneurs.
- La perte moyenne déjà calculé selon notre estimation de pertes d'opportunités.

Cependant, nous avons obtenu les visualisations graphiques suivantes :

- Un tableau regroupant les TOP 10 clients qui génèrent le plus de conteneurs bloqués.

- Deux diagrammes à bandes verticales : l'un illustrant le % des conteneurs par clusters et l'autre le % des clients par clusters (Green ,Red et Yellow).
- Deux diagrammes à bandes horizontales : l'un illustrant le nombre de conteneurs retenus par catégorie (A, B et C) et l'autre le nombres de conteneurs par type de conteneurs.
- Une carte géographique des cinq ports opérationnels de Maersk Algérie.
- Un digramme camembert illustrant le % des conteneurs retenus par intervalles de jours.

Notant que les indicateurs précédents on été validé au fure à mesure par la responsable principale des conteneurs en souffrance, ainsi que les membres de l'équipe CX à travers des séances de travail organisées en continu pour rescencer les différentes informations nécessaires à afficher dans le TDB.

Étape 6 : Collecte des données

Pour des raisons de confidentialités, les accès auxquelles nous dispositions à l'entreprise ne révèlent pas de détails sur l'axe financier du processus de gestion des conteneurs.

Cependant, les indicateurs choisis sont conçus à partir des rapports EXCEL extraits à partir d'un système d'information dédié à l'entreprise déjà expliqué dans la partie du mode opératoire.

Étape 7 : Le système de tableau de bord

Le tableau de bord est réalisé à une fréquence mensuelle pour visualiser les KPIs souhaités car les données sont rarement introduites journalièrement. Ci-dessous quelques exemples où le recours au tableau de bord permet d'anticiper certaines situations :

1. Les digrammes à barres verticales des % clients et % conteneurs par clusters facilitent la gestion du processus à travers trois groupes possédant des critères différents, le cluster Green de conteneurs par exemple se caractérise souvent par des problèmes liés à la documentation, et un nombre de jours assez bas donc pour ce clusters l'entreprise tentera d'accompagner les clients pour résoudre leurs problèmes d'échange documentaire.
2. Le tableau des TOP 10 clients permettra à l'équipe CX de : repérer avec précision les ID-Clients à contacter en premier lieu, d'avoir d'autres informations complémentaires pour analyser les différents cas de figures. Citons à titre d'exemple l'une des situations très courantes : Si le client concerné est la cause d'une rétention répétitive, l'entreprise doit intervenir immédiatement et préalablement blacklisté ce client . Ceci peut être rapidement détecté grâce à la colonne " COUNT OF BL" soit "le nombre de connaissements".

La phase de mise en oeuvre

Étape 8 : Le choix des progiciels et outils

Nous avons choisi de développer le tableau de bord sur « Power BI » car, en plus d'être celui que nous maitrisons le mieux, il offre un accès gratuit à son application de bureau"

Power BI Desktop”.

il permet de se connecter à pratiquement n’importe quelle source de données, travailler avec des volumétries considérables liées au Big Data et de créer des visualisations de données personnalisées et interactives avec une interface suffisamment simple pour que les utilisateurs finaux créent leurs propres rapports et tableaux de bord.

Nous avons remarqué aussi que ce dernier est compris dans le package hardware de l’entreprise ,mais non exploité par l’ensembles des employés.

Étape 9 : Intégration et déploiement

Le choix de ”Power BI” pour le déploiement de notre tableau de bord s’avère judicieux du fait que ce dernier est déjà intégré sur l’ensembles des bureaux.

Par conséquent, notre tableau de bord sera facilement exploité et ne nécessiter aucune formation pour la responsable CS ainsi que son équipe.

La phase de suivi permanent

Étape 10 : Audit du système Avec l’évolution de l’entreprise, la diminution de la pertinence des indicateurs peut avoir lieu. Cette diminution est traduite par un écart entre ce que porte l’indicateur comme information et la réalité sur le terrain.

Donc, il est important de conduire des audits à fin d’avoir une cohérence avec les nouveaux objectifs de l’entreprise.

Ci-dessous l’inteface Power BI du tableau de bord mis en place selon la base de données de mars 2021:

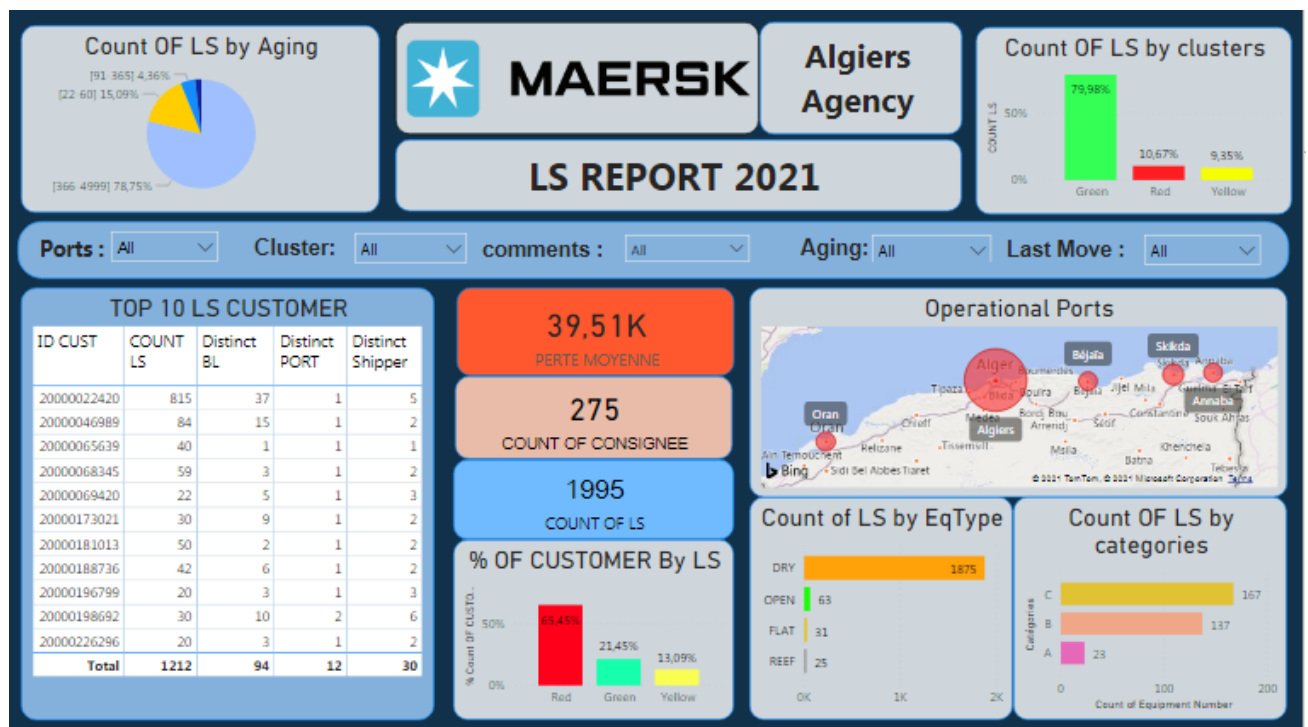


FIG. 69 : L’interface Power BI du tableau de bord élaboré

L'interface conçue regroupe l'ensemble des visualisations nécessaires à la gestion des conteneurs en souffrance (diagrammes, chiffres et filtres.) Cette dernière pourrait être lue par différents intervenants, et sera utilisée pour prendre des décisions qui permettront un gain en temps considérable à l'équipe CX.

La répartition en pourcentage des conteneurs par plages de jours (22-60 , 61-90, 91-365, 366-499) à travers le diagramme camembert présent dans le TDB précédent, favorisera le positionnement de l'équipe CX pour l'application de la SOP déjà expliquée dans le chapitre 2.

L'équipe gérante des conteneurs en souffrance chez Maersk Algérie doit donc interpréter les données disponibles dans le but d'évaluer les résultats des actions mises en place et les réviser si nécessaire.

4.6 Synthèse

Dans ce chapitre final, nous avons détaillé le processus de conception du modèle de clustering en suivant les étapes de la méthodologie de gestion de projet Machine Learning appelée CRISP-DM. En marge de cette dernière, nous avons commencé par comprendre l'objectif de notre cas, par la suite on a pris le temps de bien comprendre données collectées afin de pouvoir les préparer, modéliser et évaluer dans le but d'obtenir le modèle de clustering le plus performants.

Nous avons obtenu des modèles des résultats satisfaisants avec un coefficient de silhouette égal à 0.82 pour les conteneurs et 0.68 pour les clients ce qui représente une forte structure de clustering.

Au niveau de la deuxième partie, nous avons déroulé la démarche GIMSI avec ces 4 phases qui nous ont permis d'élaborer un tableau de bord permettant la visualisation de la performance en tenant compte des clusters formés lors de la première partie.

Conclusion

Le présent travail s'inscrit dans le cadre de l'optimisation du processus de gestion des conteneurs en souffrance. Nous avons opté pour une solution axée sur l'apprentissage non-supervisé qui tiendra compte de l'ensemble des variables caractérisant les conteneurs en souffrance pour effectuer une double segmentation, sur les clients et les conteneurs qu'ils retiennent.

La gestion des conteneurs en souffrance a toujours été un grand défi pour toute compagnie maritime voulant se distinguer des concurrents. En effet, un dysfonctionnement à ce niveau pourrait générer des pertes sur chiffres d'affaires ce qui ralentirait sa croissance, sa performance et menacerait la pérennité de son activité. C'est dans ce contexte que nous avons effectué notre étude au sein de la compagnie maritime danoise Maersk, dans le département customer experience chargé de la gestion des conteneurs en souffrance. Ces derniers représentent le coeur de notre travail et engendrent une perte de chiffre d'affaires estimée à 65 900 552 €. Afin d'apporter des réponses à notre problématique et aux hypothèses déjà formulées, il a été convenu de suivre les étapes suivantes :

- Une modélisation de l'ensemble des processus existants du département CX à l'issue des réunions organisées avec les responsables des différents services.
- Un diagnostic détaillé de la problématique par la démarche QQQQCCP, qui nous a permis d'avoir une large portée sur les actions déjà mises en place mais aussi d'identifier des pistes d'amélioration.
- Une collecte de données à travers des extractions faites sur le système de tracking des conteneurs de Maersk et en utilisant des rapports annuels et des documents internes.
- Le développement d'un modèle de clustering permettant la segmentation des conteneurs et des clients suivant la démarche CRISP-DM.
- L'implémentation d'un tableau de bord tenant compte des nouveaux segments de conteneurs et de clients à travers la démarche GIMSI. Cet outil implémenté sur Power BI permettra de visualiser les indicateurs de performance adéquats au moment souhaité et servira de support pour piloter la performance.

Les modèles d'apprentissage non-supervisé que nous avons développé ont donné des résultats très satisfaisant avec un coefficient de silhouette égal à 0.82 pour le conteneurs et 0.69 pour les clients, nous avons donc obtenu une forte structure qui peut être considérée comme référence dans la prise de décision.

La solution proposée a été validée avec succès par la responsable des agences et de la gestion des conteneurs en souffrance et a été considérée comme un pas significatif vers nouvelle approche d'amélioration des outils de gestion en incluant les modèles d'apprentissages automatiques et les améliorer d'une manière incrémentale afin de les rendre plus performants.

De ce qui précède, nous pouvons confirmer les hypothèses formulées :

- Le diagnostic effectué nous a montré que Maersk ne possède la totalité des outils permettant une visibilité sur les cas de conteneurs en souffrances.
- Les résultats de l'évaluation des modèles de clustering montre que les techniques de machine learning, apprentissage non-supervisé ainsi que les tableaux de bord peuvent être appliquées dans ce contexte.

Nous pouvons affirmer que ce travail était une excellente opportunité pour nous dans le sens où cela nous a permis de découvrir le secteur du transport maritime. Ce projet nous a permis ainsi de capitaliser les connaissances acquises durant notre cursus universitaire et d'apprendre les rouages du monde professionnel.

Pour conclure notre travail, nous posons la problématique suivante qui permettra à Maersk de s'inscrire dans une démarche d'amélioration continue vis-à-vis du modèle d'apprentissage que nous avons mis en place :

”Comment pourrait-on améliorer la performance des modèles d'apprentissage non-supervisé en tenant compte de la variable du temps ?”

Bibliographie

- ALIJANI, Shahrzad. Stratégie de gestion des retours des conteneurs maritimes de l'Internet physique. 2016. Thèse de doctorat. École Polytechnique de Montréal.
- BONACCORSO, Giuseppe. Machine learning algorithms. Packt Publishing Ltd, 2017.
- CHAPMAN, Pete, CLINTON, Julian, KERBER, Randy, et al. The CRISP-DM user guide. In : 4th CRISP-DM SIG Workshop in Brussels in March. 1999.
- Charreaux, Gérard. (1998). Le rôle de la confiance dans le système de gouvernance des entreprises. Économies et Société, série Sciences de Gestion. 8–9.
- Cornuéjols, Antoine and Miclet, Laurent. 2003. Apprentissage artificiel. s.l. : Eyrolles, 2010, 2003. p. 803. ISBN :782212124712.
- COUTANSAIS, Cyrille P. Transport maritime. Études, 2010, vol. 412, no 3, p. 307-318.
- DALKEY, Norman C. Delphi. Routledge, 2018.
- FEMINIER. B D. BOIX. (2003). « Le tableau de bord facile : Manager d'équipe », Éd d'Organisation, Paris, 2003
- FERNANDEZ, Alain. Les nouveaux tableaux de bord des décideurs. Les éditions de l'organisation,, 2000.
- Forgy, E.W. 1965. Cluster analysis of multivariate data : efficiency versus interpretability of classifications. 1965, pp. 768–769.
- FRÉMONT, Antoine et SOPPÉ, Martin. Transport maritime conteneurisé et mondialisation. In : Annales de géographie. Armand Colin, 2005. p. 187-200.
- Gacôgne, Louis. 2015. Intelligence artificielle Cours, exercices et projets. s.l. : Ellipses, 2015.
- Géron, Aurélien. (2017). Machine Learning avec Scikit-Learn. s.l. : Dunod, 2017. p. 256 .
- Glasson-Cicognani, Mélanie Berchtold, André. (2010). Imputation des données manquantes : Comparaison de différentes approches. 42èmes Journées de Statistique.

REINHARDT, Forest L., CASADESUS-MASANELL, Ramon, et NELLEMAN, Frederik. Maersk Line and the Future of Container Shipping. Harvard Business School Case Study, 2012, p. 9-712.

Planchon, Viviane. (2005). Traitement des valeurs aberrantes : concepts actuels et tendances générales. Biotechnologie, Agronomie, Société et Environnement. 9.

SIRIMANNE, Shamika N., HOFFMAN, J., JUAN, W., et al. Review of maritime transport, 2019. tech. rep, 2020.

V. Barnett and T. Lewis. (1994). Outliers in statistical data 3rd edition, (John Wiley Sons, Chichester), 584 pp., [UK pound]55.00, ISBN 0-471-93094-6

VOYER, Pierre. Tableaux de bord de gestion et indicateurs de performance : 2e édition. PUQ, 2011.

XU, Rui et WUNSCH Don. (2008). Clustering. John Wiley Sons.

ZERKAK, Sabrina Encadreur, AISSANI, Saloua, et YAHI, Dalila. Le tableau de bord outil de contrôle de gestion. 2013. Thèse de doctorat. Université de bejaia.

Webographie

Documentation officielle de la bibliothèque scikit-learn. Disponible sur : <https://scikit-learn.org/>

Henri Michel. (2019). Google Colab : Le guide Ultime. Disponible sur : <https://ledatascientist.com/google-colab-le-guide-ultime/>

Kamil Krzyk. (2018). Coding Deep Learning For Beginners. Disponible sur : <https://towardsdatascience.com/coding-deep-learning-for-beginners-types-of-machine-learning-b9e651e1ed9d>

Khalil Elleuch. (2021). Introduction à la réduction de dimensionnalité en Machine Learning. Disponible sur : <https://www.invivoo.com/reduction-dimensionnalite-machine-learning/>

Maersk investor presentation. (2019). Disponible sur : <https://investor.maersk.com/static-files/5f4e7940-a5bd-4e23-89ce-686ea682bf12>

Manimaran. (2019). Clustering Evaluation strategies. Disponible sur : <https://towardsdatascience.com/clustering-evaluation-strategies-98a4006fcfc#:~:text=Clustering%20is%20an%20unsupervised%20machine%20learning%20algorithm.,not%20contain%20ground%20truth%20labels.>

MIAILHE, Nicolas et HODES, Cyrus. (2018) La troisième ère de l'intelligence artificielle. Disponible sur : https://www.institut.veolia.org/sites/g/files/dvc2551/files/document/2018/03/Facts-AI-03_La_troisieme_ere_de_lintelligence_artificielle_-_Nicolas_Miailhe_Cyrus_Hodes

Annexes

Annexe A

Modélisation du processus d'export

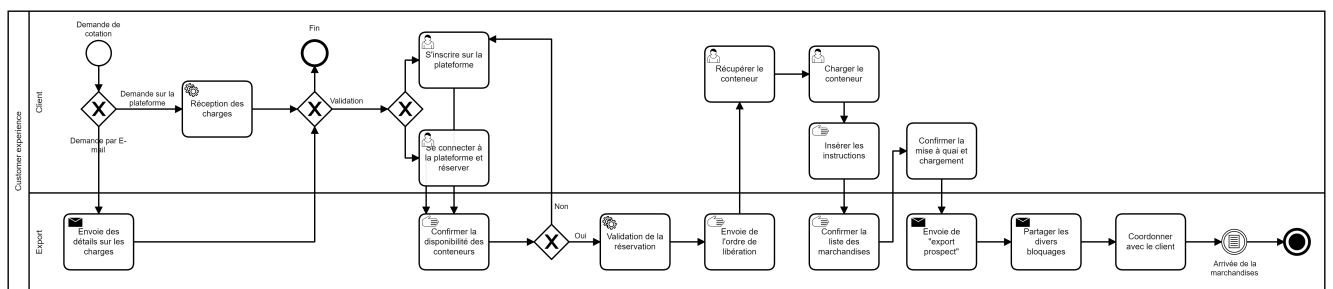


FIG. 70 : Modélisation BPMN du processus d'export

Annexe B

Modélisation du processus d'import

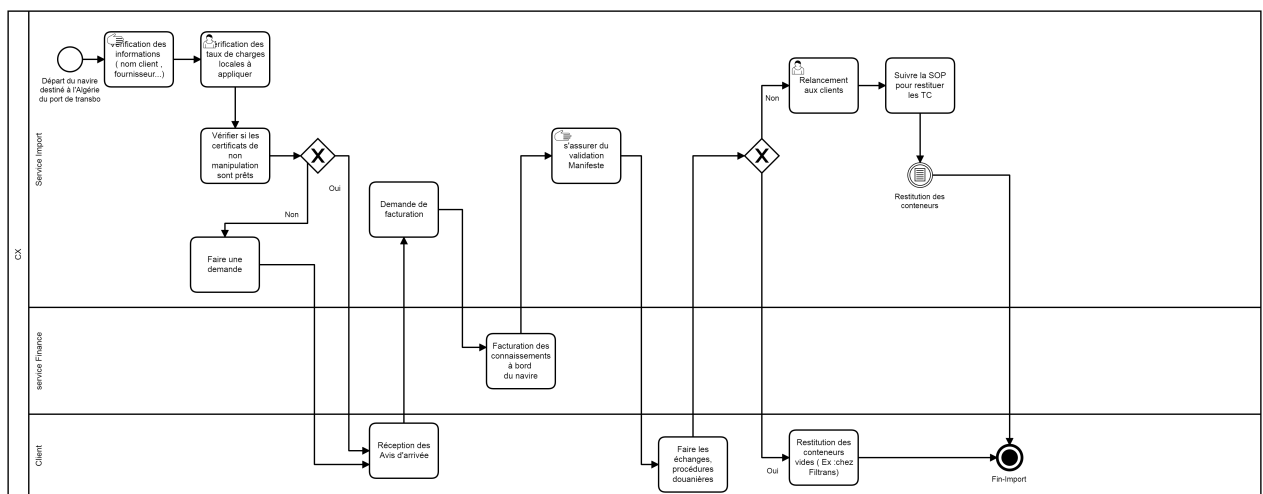


FIG. 71 : Modélisation BPMN du processus d'import

Annexe C

Modélisation du processus de traitement des réclamations

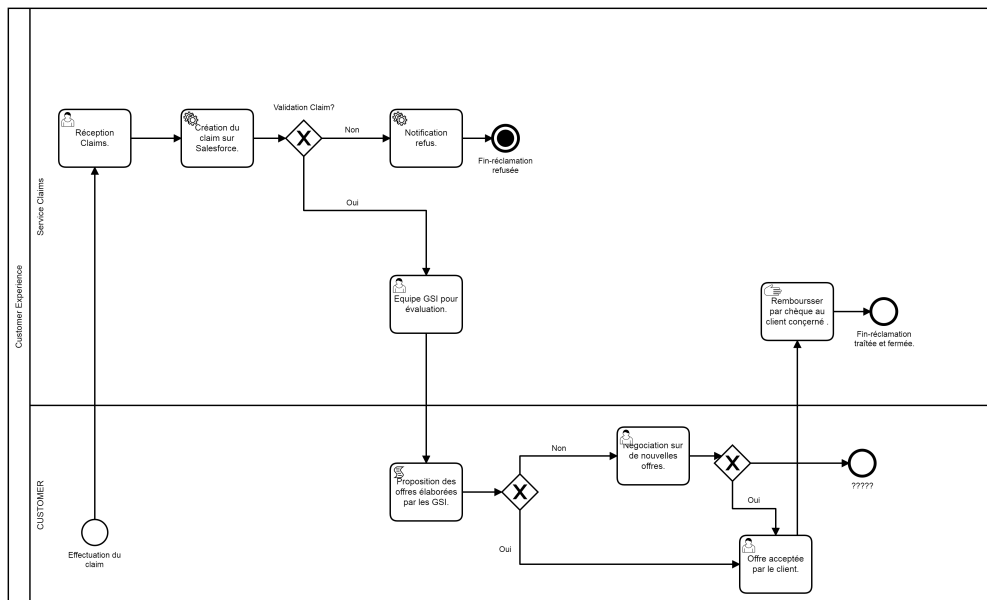


FIG. 72 : Modélisation BPMN du processus d'import

Annexe D

Modélisation du processus d'élaboration du manifest

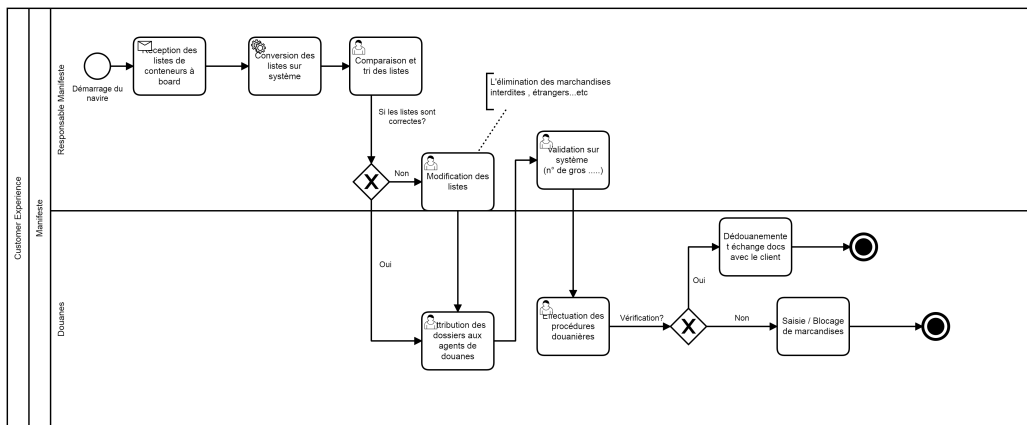


FIG. 73 : Modélisation BPMN du processus d'élaboration du manifest

Annexe E

Modélisation du processus d'échange documentaire

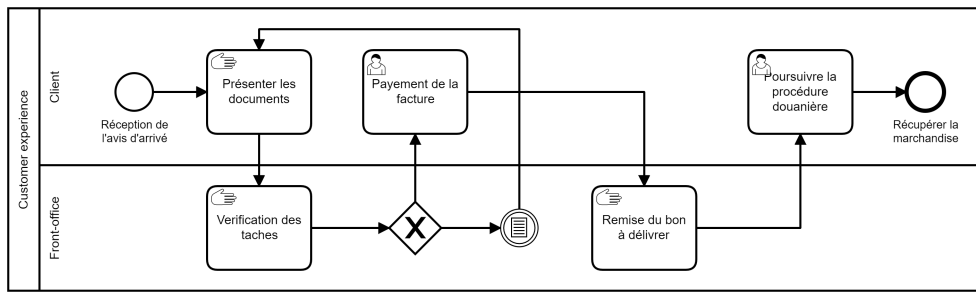


FIG. 74 : Modélisation BPMN du processus d'échange documentaire

Annexe F

Demande de cotation

Offer 1-1				
Scheduled Route	Algiers Terminal, Algeria --> Tanger Med 2, Morocco --> Havre (Le) / Cnm Terminal, Seine-Maritime, France			
Mode of Transport	Ocean > Ocean			
Place of Receipt	Algiers port, Algeria	Rate Validity	From 28-Apr-2021 to 30-Apr-2021	
Place of Delivery	Le Havre, Seine-Maritime, France	Service Mode	CY / CY	
Last Acceptance Date	28-Apr-2021	Commodity	FAK	
Transit Time	31 day(s)	Quoted for weight	18.0 TONS	
Surcharge Name	Basis	Currency	40FLAT	Surcharge Type
Basic Ocean Freight (Fixed)	Container	EUR	900	Freight
Congestion Fee Origin (floating as per tariff)	Container	USD	0	Freight
Peak Season Surcharge (floating as per tariff)	Container	USD	0	Freight
Special Equipment Service (floating as per tariff)	Container	EUR	800	Freight
Congestion Fee Destination (floating as per tariff)	Container	USD	0	Freight
Winter Fee (floating as per tariff)	Container	EUR	0	Freight
Environmental Fuel Fee (floating as per tariff)	Container	EUR	100	Freight
Free In Service (floating as per tariff)	Bill of Lading	DZD	8200	Origin
Terminal Handling Service - Origin (floating as per tariff)	Container	DZD	16000	Origin
Algerian Stamp Origin (floating as per tariff)	Bill of Lading	DZD	1000	Origin
Equipment Positioning Service - Export (CY) (floating as per tariff)	Container	USD	0	Origin
Import Service (floating as per tariff)	Container	EUR	230	Destination
Equipment Positioning Service - Import (CY) (floating as per tariff)	Container	USD	0	Destination

FIG. 75 : Exemple d'une demande de cotation effectuée sur maersk.com.

Annexe G

Trame des entretiens



FIG. 76 : Trame indiquant le déroulement des entretiens.

Annexe H

Scatter-matrix K-Means

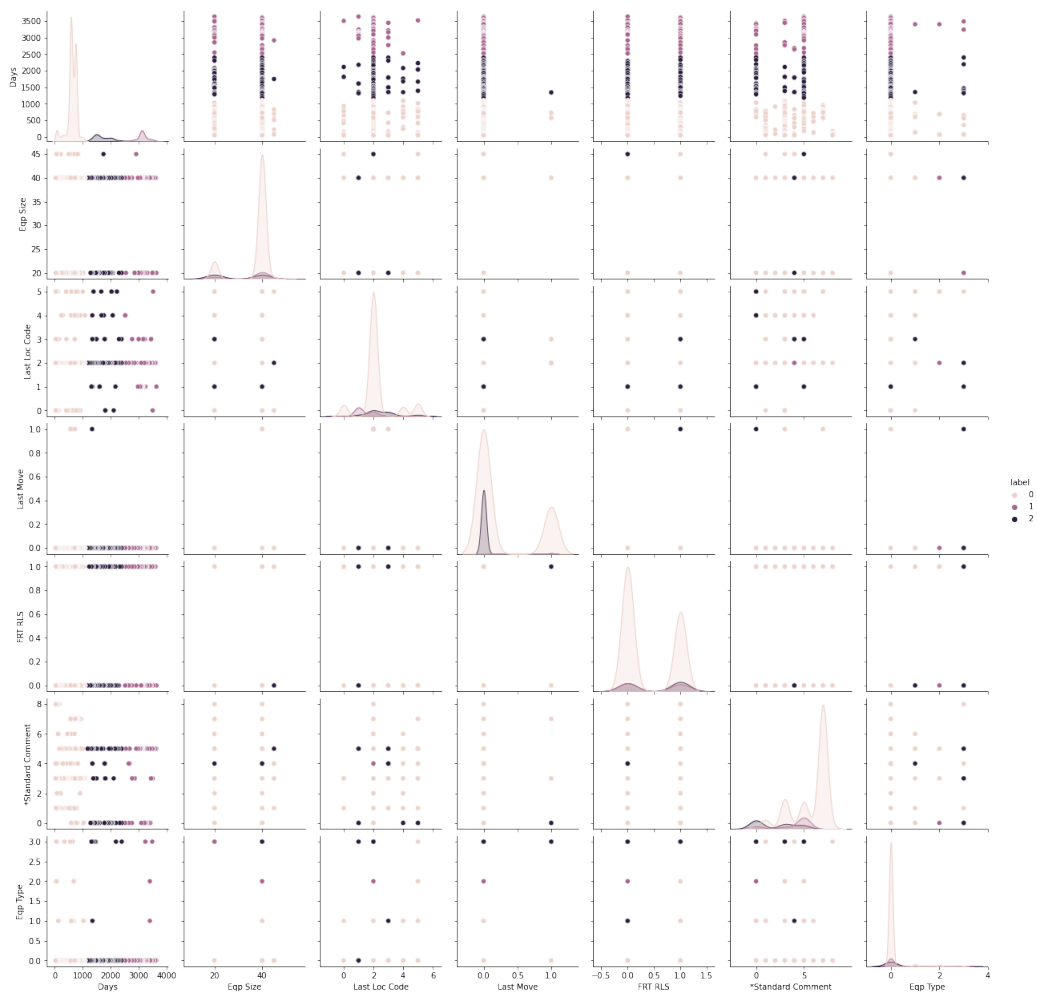


FIG. 77 : Scatter-matrix montrant les résultats du clustering obtenus par K-Means.