



Titre: Modélisation du Capital Humain dans le processus d'innovation
Title: d'une entreprise industrielle

Auteur: Thibaut Magalas
Author:

Date: 2021

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Magalas, T. (2021). Modélisation du Capital Humain dans le processus
Citation: d'innovation d'une entreprise industrielle [Mémoire de maîtrise, Polytechnique
Montréal]. PolyPublie. <https://publications.polymtl.ca/9958/>

 **Document en libre accès dans PolyPublie**
Open Access document in PolyPublie

URL de PolyPublie: <https://publications.polymtl.ca/9958/>
PolyPublie URL:

**Directeurs de
recherche:** Catherine Beaudry, & Bruno Agard
Advisors:

Programme: Maîtrise recherche en génie industriel
Program:

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

**Modélisation du Capital Humain dans le processus d'innovation d'une
entreprise industrielle**

THIBAUT MAGALAS

Département de mathématiques et de génie industriel

Mémoire présenté en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*

Génie industriel

Décembre 2021

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Ce mémoire intitulé :

Modélisation du Capital Humain dans le processus d'innovation d'une entreprise industrielle

présenté par **Thibaut MAGALAS**

en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*

a été dûment accepté par le jury d'examen constitué de :

Camélia DADOUCHI, présidente

Catherine BEAUDRY, membre et directrice de recherche

Bruno AGARD, membre et codirecteur de recherche

Virginie FRANCOEUR, membre

DÉDICACE

A toute ma famille pour la confiance qu'elle m'a accordée pendant tout ce programme à l'École Polytechnique de Montréal.

REMERCIEMENTS

En premier lieu, je remercie Catherine et Bruno pour m’avoir donné la chance de rejoindre le programme de Maîtrise Recherche au département de Mathématiques Appliquées et de Génie Industriel (MAGI) de Polytechnique Montréal, en m’accueillant respectivement à la Chaire-Innovation et au Laboratoire en Intelligence des Données.

Je remercie également les professeurs du département de MAGI qui ont partagé leur savoir et savoir-faire pour la réalisation de ce projet. Je suis également reconnaissant pour l’opportunité que j’ai eu de participer à l’école d’été HEC Montréal MOSAIC 2020 qui était une expérience très riche sur le thème de la résilience et le Management de la créativité. Je remercie également mes collègues de la Chaire et du LID qui m’ont aidé à concrétiser ce projet.

De plus, j’aimerais porter attention à Annie et Mikel pour les travaux de groupes que nous avons mené ensemble durant mon programme de Maîtrise Recherche et qui m’ont permis d’acquérir des compétences de recherche que j’ai pu mettre en œuvre lors de ce projet.

Enfin, je souhaite remercier particulièrement mes parents qui m’ont toujours soutenu dans mes études longues et qui ont toujours eu confiance en ma détermination pour atteindre mes objectifs. Finalement, je n’oublie pas ma sœur Pauline pour son soutien et sa présence à Montréal pendant la durée de mon programme et dont je suis très fier car elle est récemment diplômée de HEC Montréal.

RÉSUMÉ

Dans le monde numérique caractéristique du XXI^e siècle, une entreprise qui souhaite se démarquer dans son industrie doit proposer et déployer de nouvelles innovations technologiques. Une entreprise au cœur de l'innovation technologique est d'autant plus performante qu'elle excelle dans la gestion de la connaissance. Il s'agit alors pour une entreprise industrielle avant tout de mieux comprendre les interactions entre la connaissance aussi bien individuelle qu'organisationnelle (interne) et le processus d'innovation. Le processus d'innovation en entreprise industrielle était jusqu'alors considéré comme linéaire, ce n'est que récemment que le processus d'innovation dans une entreprise est envisagé comme collaboratif et itératif. Très peu d'études dans la littérature ont tenté de modéliser le rôle de l'individu dans le processus d'innovation d'une entreprise industrielle. C'est donc à partir de ce nouveau cadre conceptuel du processus d'innovation ainsi que d'un état de l'art consistant sur le rôle du Capital Intellectuel d'une entreprise dans le processus d'innovation que nous construisons par le biais d'une étude exploratoire 3 modèles qui impliquent le Capital Humain des employés et le potentiel d'innovation de l'entreprise industrielle. L'objectif de ce mémoire est alors d'identifier le potentiel d'innovation du Capital Organisationnel (CO) d'une entreprise industrielle à partir du Capital Humain (CH) de ses employés. Cette étude exploratoire est l'une des premières dans la littérature à mettre en œuvre des méthodes d'apprentissage d'extraction de connaissances à partir de la donnée, afin de proposer un outil de mécanisme de soutien au processus d'innovation des entreprises industrielles en ce qui a trait au Capital Humain de sa force de travail interne. Pour ce faire, nous réalisons l'étude exploratoire sur un ensemble de données de l'entreprise CAE Inc., qui est une entreprise basée au Canada qui œuvre dans les secteurs industriels de l'aéronautique et de la santé. D'une part, nous collectons les données du Capital Humain des employés à partir de la source de données LinkedIn avec respect de la confidentialité des données et des données personnelles, et d'autre part nous collectons les données du potentiel d'innovation de l'entreprise avec les données des brevets de CAE Inc. issues de la plateforme Patseer. Nous obtenons alors un ensemble total composé de 493 individus (employés) de l'entreprise CAE Inc. Lors de cette étude exploratoire, nous développons un outil de moissonnage de données qui permet l'automatisation de la collecte et du traitement des données du Capital Humain. Afin de réaliser l'objectif de cette étude exploratoire, les méthodes d'analyse de données qui sont mises en œuvre sont les règles d'association, les arbres de décision et un algorithme de classification. Les analyses effectuées amènent à la construction d'un descripteur

nommé potentiel d'innovation qui permet l'identification en 5 classes de tendances d'innovation pour chaque employé. Ainsi les résultats des analyses sont novateurs pour identifier le lien entre les dimensions de savoir-faire, du développement du Capital Humain des employés et le potentiel d'innovation du Capital Organisationnel de l'entreprise industrielle par la publication de brevets. Néanmoins, l'étude exploratoire présente plusieurs limites qui sont notamment l'utilisation directe de ces résultats par l'entreprise car les potentiels d'innovation obtenus pour les employés sont génériques et ne présentent pas de particularités spécifiques à des projets ou des activités au niveau de l'équipe. De plus, les données du Capital Humain proviennent du réseau professionnel LinkedIn, ces données peuvent présenter un biais de non-conformité et de désirabilité sociale. Toutefois les résultats obtenus par cette construction de modèles et les analyses suggèrent qu'avec l'utilisation d'un modèle enrichi par des indicateurs supplémentaires sur d'autres dimensions du Capital Humain, l'accès à d'autres indicateurs pour le potentiel d'innovation et d'un ensemble avec davantage d'individus il serait sûrement possible de construire une modélisation plus fine de l'identification du potentiel d'innovation du Capital Organisationnel d'une entreprise industrielle spécifique à une équipe ou à un projet à partir du Capital Humain.

Mots-clés : indicateurs d'innovation, constitution d'équipe, capital humain, capital intellectuel, exploration de données, LinkedIn

ABSTRACT

In the digital world of the 21st century, a company that wants to stand out in its industry must propose and deploy new technological innovations. A company at the heart of technological innovation is all the more successful if it excels in knowledge management. For an industrial company, it is above all a question of better understanding the interactions between individual and organizational (internal) knowledge and the innovation process. The innovation process in an industrial company was previously considered as linear; it is only recently that the innovation process in a company is considered as collaborative and iterative. Very few studies in the literature have attempted to model the role of the individual in the innovation process of an industrial company. It is therefore from this new conceptual framework of the innovation process as well as a consistent state of the art on the role of the intellectual capital of a company in the innovation process that we build through an exploratory study 3 models that involve the human capital of employees and the innovation potential of the industrial company. The objective of this dissertation is then to identify the innovation potential of the Organizational Capital (OC) of an industrial firm from the Human Capital (HC) of its employees. This exploratory study is one of the first in the literature to implement knowledge extraction learning methods from data, in order to propose a support mechanism tool for the innovation process of industrial companies with respect to the Human Capital of its internal workforce. To do this, we conduct an exploratory study on a sample of data from CAE Inc, a Canadian-based company that operates in the aeronautics and health industrial sectors. On the one hand, we collect the Human Capital data of the employees from the LinkedIn data source with respect to data confidentiality and personal data, and on the other hand we collect the data of the innovation potential of the company with the patent data of CAE Inc. from the Patseer platform. We then obtain a total sample of 493 individuals (employees) from the company CAE Inc. During this exploratory study, we develop a data harvesting tool that allows the automation of the collection and processing of Human Capital data. In order to achieve the objective of this exploratory study, the data analysis methods that are implemented are association rules, decision trees and a classification algorithm. The analyses carried out lead to the construction of a descriptor named innovation potential which allows the identification of 5 classes of innovation tendencies for each employee. Thus, the results of the analyses are innovative in identifying the link between the dimensions of know-how, of the development of the employees' Human Capital and the innovation potential of the Organizational Capital of the industrial company

through the publication of patents. Nevertheless, the exploratory study has several limitations which include the direct use of these results by the company because the innovation potentials obtained for the employees are generic and do not present specificities of projects or activities at the team level. In addition, the Human Capital data comes from the LinkedIn professional network, this data may have a non-conformity and social desirability bias. However, the results obtained from this model building and analysis suggest that with the use of a model enriched with additional indicators on other dimensions of Human Capital, access to other indicators for innovation potential, and a sample with more individuals it would surely be possible to build a finer modeling of the identification of the innovation potential of the Organizational Capital of an industrial company specific to a team or a project from Human Capital.

Keywords: Innovation indicators, team constitution, human capital, intellectual capital, data mining, LinkedIn

TABLE DES MATIÈRES

DÉDICACE.....	III
REMERCIEMENTS	IV
RÉSUMÉ.....	V
ABSTRACT.....	VII
TABLE DES MATIÈRES	IX
LISTE DES TABLEAUX.....	XII
LISTE DES FIGURES.....	XIII
LISTE DES SIGLES ET ABRÉVIATIONS	XV
CHAPITRE 1 INTRODUCTION.....	1
CHAPITRE 2 REVUE DE LITTÉRATURE	4
2.1 L’innovation.....	4
2.1.1 Les investissements pour l’innovation technologique	5
2.1.2 Les modélisations récentes des processus d’innovation	6
2.2 Le Capital Humain	8
2.2.1 Le rôle du Capital Humain (CH) dans le processus d’innovation.....	9
2.2.2 Les indicateurs du Capital Humain (CH).....	12
2.3 Les sources de données du Capital Humain en entreprise	17
2.3.1 Le réseau professionnel LinkedIn	18
2.3.2 Les compétences du Capital Humain à l’ère du numérique.....	19
CHAPITRE 3 CADRE DE LA RECHERCHE.....	23
3.1 Contexte et question de recherche.....	23
3.1.1 Contexte et question de recherche du projet	23
3.1.2 Objectifs	24

3.2	Contribution originale	25
3.3	Cadre conceptuel et modèles	26
CHAPITRE 4 MÉTHODOLOGIE		30
4.1	Données	30
4.1.1	Extraction d'information des données brevet.....	31
4.1.2	Extraction d'information des données LinkedIn	35
4.1.3	Résumé de la sélection des informations	36
4.2	Sélection des données.....	37
4.2.1	Sélection des données d'entraînement	38
4.2.2	Sélection des données de test	40
4.3	Traitement des données	43
4.4	Construction des variables	46
4.4.1	Construction des variables de données de brevets	46
4.4.2	Construction des variables de données LinkedIn	49
4.5	Modèle d'analyses.....	55
4.5.1	Règles d'associations	56
4.5.2	Arbres de décision	59
4.5.3	Algorithme des K-Moyennes	62
CHAPITRE 5 RÉSULTATS		64
5.1	Statistiques descriptives	64
5.2	Analyses statistiques	69
5.2.1	Règles d'association.....	69
5.2.2	Arbre de décision du Modèle 2	72
5.2.3	Algorithme des K-Moyenne du Modèle 3.....	73

CHAPITRE 6	DISCUSSION, CONCLUSION ET RECOMMANDATIONS.....	81
6.1	Interprétation des résultats	81
6.2	Discussion	85
6.2.1	Évaluation de la qualité et de la validité de l'étude.....	85
6.2.2	Recommandations	86
6.3	Conclusion et limites de la recherche.....	87
6.3.1	Conclusion.....	87
6.3.2	Limites de la recherche	87
RÉFÉRENCES	89

LISTE DES TABLEAUX

Tableau 2-1. Modèles théoriques sur le CI en entreprise	10
Tableau 2-2. Structure du concept du CH	16
Tableau 4-1. Liste des attributs des données brevets collectés sur Patseer	32
Tableau 4-3. Résumé de la sélection des variables.	37
Tableau 4-2. Résultats de la recherche des profils sur LinkedIn	41
Tableau 4-4. Construction des catégories de "sous-domaines technologiques" des brevets.....	47
Tableau 4-5. Catégories des sous-domaines technologiques par domaines technologiques des brevets	48
Tableau 4-6. Construction des variables binaires de Domaine d'études	51
Tableau 4-7. Construction de la variable catégorique "Domaine d'études"	52
Tableau 4-8. Construction de la variable catégorique "Diversité des domaines d'études"	53
Tableau 4-9. Construction de la variable "Catégorie de compétences"	54
Tableau 4-10. Construction de la variable catégorique "Diversité des catégories de compétences"	55
Tableau 5-1. Résultats des règles d'association à partir du Modèle 1 de variables binaires	70
Tableau 5-2. Analyse des descripteurs du potentiel d'innovation pour chacune des classes issues de l'algorithme des K-Moyennes.....	75
Tableau 6-1. Hypothèses et résultats obtenus avec l'algorithme de classification sur le Modèle 3	84

LISTE DES FIGURES

Figure 2-1 Arborescence des types d'innovation, Interprété par Beaudry C. de (OCDE/Eurostat, 2019).....	5
Figure 2-2 Structure interne du <i>Middle-ground</i> dans le milieu organisationnel. Adapté de (Cohendet & Simon, 2020)	7
Figure 3-1. Modèle 1 de variables binaires	28
Figure 3-2. Modèle 2 de variables binaires et discrètes	28
Figure 3-3. Modèle 3 de variables binaires, discrètes et catégoriques	29
Figure 4-1. Étapes de sélection de l'ensemble d'entraînement.....	39
Figure 4-2. Traitement de données LinkedIn	45
Figure 4-3. Exemple d'arbre de décision de Classification Ascendante Hiérarchique.....	61
Figure 5-1. Distribution des tranches d'âge parmi la population	64
Figure 5-2. Distribution des établissements académiques fréquentés par la population.....	65
Figure 5-3. Distribution des catégories de domaines d'études	65
Figure 5-4. Distribution des catégories de diversité des domaines d'études parmi la population..	66
Figure 5-5. Distribution des niveaux d'études parmi la population	66
Figure 5-6. Distribution des groupes de compétences à travers la population.....	67
Figure 5-7. Distribution des catégories de diversité de groupes de compétence parmi la population	68
Figure 5-8. Distribution du nombre de brevets réalisés par la population	68
Figure 5-9. Dendrogramme (seuil = 5.0 classes)	72
Figure 5-10. Graphique de l'Analyse en Composantes Multiples (ACM) à 2 facteurs sur les 3 variables catégoriques du Modèle 3	74

Figure 5-11. Graphe de la fonction coût de l'algorithme des K-Moyennes sur le Modèle 3. Entropie en fonction du nombre de classes K.....	75
Figure 5-12. Distribution des profils par classe selon leurs domaines d'études	77
Figure 5-13. Distribution des profils par classe selon la Diversité des groupes de compétences ..	78
Figure 5-14. Distribution des profils par classe selon le niveau d'études	79
Figure 5-15. Distribution des profils par classe selon la diversité des études.....	80

LISTE DES SIGLES ET ABRÉVIATIONS

La liste des sigles et abréviations présente, dans l'ordre alphabétique, les sigles et abréviations utilisés dans le mémoire ou la thèse ainsi que leur signification. En voici quelques exemples :

4POINT0	Partenariat pour l'Organisation de l'Innovation et des Nouvelles Technologies
BMF	<i>Business- Management - Finance</i>
CH	Capital Humain
CI	Capital Intellectuel
CITP	Classification Internationale Type des Professions
CO	Capital Organisationnel
CS	Capital Social
ECD	Extraction de Connaissance à partir de Données
FEM	Forum de l'Économie Mondiale
GHCR	<i>Global Human Capital Report</i>
IPC	<i>International Patent Classification</i>
OIT	Organisation Internationale du Travail
OCDE	Organisation de Coopération et de Développement Économique
STI	Sciences et Techniques Industrielles
TIC	Technologies de l'Information et de la Communication
WIPO	<i>World Intellectual Property Office</i>

CHAPITRE 1 INTRODUCTION

« Innovation starts with people, making the human capital within the workforce decisive. »

(Anderson, 2008)

La transformation numérique des entreprises joue un rôle majeur dans la croissance économique (Xu, 2014). Dans le monde numérique caractéristique du XXI^e siècle, une entreprise qui souhaite se démarquer dans son industrie doit proposer et déployer de nouvelles innovations technologiques (Summa, 2016 ; Li, 2018). Ces changements technologiques dans les entreprises causent également des changements dans les structures de management qui deviennent de plus en plus verticales avec des prises de décision décentralisées et appuyées par l'automatisation (Van Laar et al., 2017). Dans le *Manuel d'Oslo* de 2018, l'OCDE (Organisation de Coopération et de Développement Économique) révèle les besoins créés par le développement technologique rapide durant les dernières décennies et précise notamment que la connaissance dans les entreprises est un facteur majeur pour une croissance durable d'une économie basée sur l'innovation. Une entreprise au cœur de l'innovation qui souhaite créer de la valeur doit d'abord parvenir à créer et gérer des connaissances (Madhavan & Grover, 1998). Dans le domaine de gestion des connaissances, la connaissance correspond à l'appropriation et l'interprétation des informations par des êtres humains. Il existe deux grands types de connaissances à savoir les connaissances explicites et les connaissances tacites. Les connaissances explicites sont les connaissances transcrites dans un langage formel, elles apparaissent sous forme tangible comme un document manuscrit ou électronique. Les connaissances tacites désignent les compétences innées ou acquises, le savoir-faire et l'expérience, elles sont généralement plus difficiles à formaliser que les connaissances explicites et par conséquent plus difficiles à retranscrire et à transmettre (Nonaka & von Krogh, 2009). En entreprise, on peut différencier les connaissances internes qui sont liées aux produits et services, aux compétences des employés ou encore aux méthodes de leadership ; et les connaissances externes comme la connaissance du marché, des concurrents, des caractéristiques des clients et des tendances technologiques. Dans la suite de ce mémoire, lorsque le terme connaissance est utilisé il s'agit des connaissances internes de l'entreprise. À ce propos, des études sur les relations de causalité entre la connaissance et l'innovation en entreprise sont

réalisées depuis plusieurs années (p. ex. : Ahuja, 2000 ; Dougherty, 1992 ; Subramaniam & Venkatraman, 2001 ; Tsai & Ghoshal, 1998). La nature précise de la connaissance, explicite ou tacite, et son utilisation par une organisation pour créer l'innovation restent encore peu caractérisées et méritent d'être approfondies (Subramaniam & Youndt, 2005). Quels sont les types de connaissances dans le milieu organisationnel ? Pour un changement de structure organisationnelle, quels sont les indicateurs liés aux types de connaissances qui maximisent le potentiel d'innovation d'une entreprise ? La capacité d'une organisation à créer et gérer des connaissances est nommée Capital Intellectuel (CI). Le Capital Intellectuel s'explique comme la somme des connaissances utilisées par une organisation pour créer un avantage compétitif (Nahapiet & Ghoshal, 1998 ; Youndt et al., 2004 ; Subramaniam & Youndt, 2005). On retrouve trois aspects proéminents du Capital Intellectuel, communément appelés Capital Social (CS), Capital Organisationnel (CO) et Capital Humain (CH). Le Capital Social désigne les connaissances disponibles, implantées et utilisées au travers des interactions individuelles et leurs réseaux d'interrelations (Nahapiet & Ghoshal, 1998). Le Capital Organisationnel fait référence aux connaissances tacites et connaissances explicites d'une entité. Plus précisément, le Capital Organisationnel est la connaissance institutionnelle et l'expérience codifiée qui est utilisée et réside dans les bases de données, les manuels, les structures, les systèmes, les procédés ainsi que dans les brevets (Youndt et al., 2004). Finalement, le Capital Humain désigne les connaissances et les compétences que l'individu possède pour créer de la valeur dans le système économique mondial (Le Chapelain & Matéos, 2020 ; WEF, 2017). Le Capital Humain est alors le point d'ancrage du Capital Intellectuel dans le processus d'innovation. De nombreux chercheurs précisent que l'innovation commence avec les personnes (Anderson, 2008). D'après les nombreuses études référencées par van Laar et al. (2017), ce sont précisément les employés qui permettent à l'entreprise de rendre le Capital Humain un facteur décisif si ce n'est primordial dans le processus d'innovation.

L'objectif de ce mémoire est d'établir un mécanisme de soutien à l'innovation avec le développement d'un modèle d'apprentissage pour identifier le potentiel d'innovation du Capital Organisationnel (CO) d'une entreprise d'un secteur industriel à partir du Capital Humain (CH) des employés. Cet objectif entre à la fois dans une perspective théorique pour l'avancement des connaissances mais également dans une perspective pratique pour toute organisation voulant constituer une équipe la plus performante possible afin d'obtenir la meilleure innovation, ou le plus

rapidement possible. En reprenant les termes déjà évoqués, il s'agit d'établir un modèle pour identifier le potentiel d'innovation du Capital Organisationnel de l'entreprise à partir du Capital Humain des employés.

Le mémoire comporte 7 chapitres. Le chapitre 2 est réservé à une revue de littérature qui permet à la fois de développer le nouveau cadre conceptuel du processus d'innovation collaboratif et itératif d'une entreprise industrielle ainsi que d'effectuer un état de l'art sur le rôle du CI dans le processus d'innovation. Cette revue de littérature nous offre alors des pistes afin de construire nos modèles.

Ensuite, le chapitre 3 présente le cadre de la recherche incluant le contexte, les objectifs, la contribution originale ainsi que le cadre conceptuel et le modèle pour ce projet de recherche. Ce cadre de la recherche est utile afin de comprendre l'entreprise industrielle qui constitue le cas d'application de l'étude exploratoire qui suit. De plus, le cadre de la recherche permet de cerner précisément comment sont construits les différents modèles qui feront l'objet d'analyses ensuite. Cette partie est indispensable pour préciser les dimensions des concepts du CH et du potentiel d'innovation du CO qui sont recherchés respectivement à travers les sources de données LinkedIn et Patseer.

Par la suite, le chapitre 4 détaille la méthodologie de recherche. La méthodologie de recherche envisagée pour l'étude exploratoire précise la méthode de sélection pour les individus (employés) de l'entreprise CAE Inc. De plus, ce chapitre 4 permet d'expliquer le traitement des données et les outils mise en place pour y parvenir. Nous présenterons également la création des variables utilisées dans les 3 modèles étudiés. Pour terminer ce chapitre nous détaillerons les méthodes d'analyse de données appliquées à nos 3 modèles.

Le chapitre 5 présente les résultats des expérimentations. Il s'agit ici d'assurer avec rigueur les analyses statistiques par des statistiques descriptives de notre ensemble, suivies des résultats des méthodes d'apprentissages sur nos 3 modèles avec les méthodes de règles d'association, d'arbre de décisions et d'algorithme de classification. Ce chapitre sur les résultats des expérimentations assure alors la fiabilité interne de la recherche

Finalement, le chapitre 6 expose les discussions et limites de la recherche. Ce dernier chapitre offre l'interprétation des résultats et le caractère novateur de ce mémoire, ainsi qu'une discussion sur la validité et la fiabilité de la recherche. Enfin ce dernier chapitre présente également les limites de l'étude exploratoire.

CHAPITRE 2 REVUE DE LITTÉRATURE

Ce chapitre recense la littérature pertinente sur l'innovation technologique, la propriété intellectuelle en entreprise, le Capital Humain, les indicateurs pour mesurer le Capital Humain et l'utilisation des données LinkedIn. La section 2.1 de cette revue de littérature porte sur l'innovation, définit le concept et la typologie sous-jacente de l'innovation de l'OCDE (2018) utilisés à ce jour dans la littérature. Nous utiliserons ces éléments comme références dans mémoire. De plus, cette section inclue une section sur l'innovation, une section sur les investissements en entreprise pour favoriser et produire l'innovation ainsi qu'une section sur les modélisations récentes du processus d'innovation en entreprise, dont l'étude exploratoire de ce mémoire a pour but de contribuer à la compréhension de certaines étapes du processus d'innovation en entreprise dans un secteur industriel. Dans la section 2.2 de ce chapitre, nous portons un point d'attention au concept du Capital Humain, à la position du Capital Humain dans le processus d'innovation ainsi qu'à une description détaillée du cadre conceptuel sur le Capital Humain proposé par le Forum Économique Mondial - FEM (2017), dans le *Global Human Capital Report*. Ce cadre conceptuel du Capital Humain sert de référence pour définir les dimensions et les indicateurs utilisés lors de l'étude exploratoire.

2.1 L'innovation

L'OCDE définit l'innovation comme « un produit ou un processus (ou une combinaison des deux) nouveau ou amélioré qui diffère sensiblement des produits ou processus précédents d'une unité¹ et qui a été mis à la disposition d'utilisateurs potentiels (produit) ou mis en œuvre par l'unité (processus) » (OCDE, 2018, p. 20). L'innovation peut être divisée en deux grandes catégories : l'innovation d'entreprise et l'innovation sociale. L'innovation d'entreprise est elle-même subdivisée en deux types : l'innovation de processus d'affaires et l'innovation de produit (Figure 2-1). La première sous-catégorie comprend la production des biens et services ; la distribution et la logistique ; la commercialisation et la vente ; les systèmes d'information et de communication ;

¹ Le terme générique « unité » est employé pour désigner l'acteur responsable de l'innovation. Il s'agit de toute unité institutionnelle d'un secteur quel qu'il soit, y compris les ménages et chacun de leurs membres.

l'administration et la gestion ; et le développement de produits et de processus d'affaires. De son côté, l'innovation de produit intègre les biens et les services distribués aux utilisateurs. L'arbre ci-dessous représente la déclinaison des différents types d'innovation.

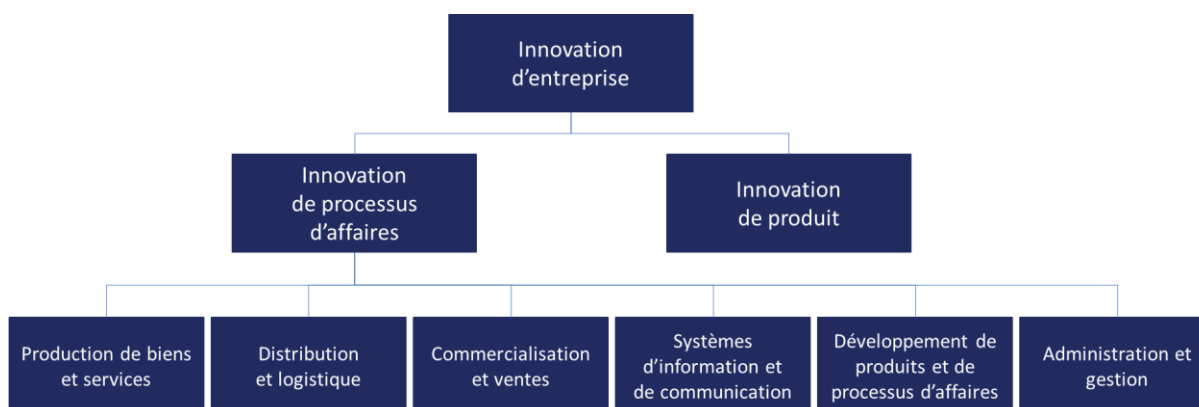


Figure 2-1 Arborescence des types d'innovation, Interprété par Beaudry C. de
(OCDE/Eurostat, 2019)

De plus, une innovation est caractérisée par son degré de nouveauté. Une innovation radicale se manifeste par des changements majeurs en rupture avec l'état passé des connaissances et s'incarnant par la création de nouveaux supports (produits, procédés, compétences, savoir-faire). À l'opposé, une innovation incrémentale est le produit de changements progressifs se matérialisant par une recombinaison ou une amélioration des supports existants (Delaplace, 2008).

2.1.1 Les investissements pour l'innovation technologique

L'innovation en entreprise est nécessaire si l'entreprise souhaite rester compétitive économiquement (Suciu & Năsulea, 2019). Plus précisément, l'innovation technologique confère des avantages cruciaux aux entreprises, comme le développement de nouveaux produits, des améliorations de qualité et de différenciation, des réductions des coûts et des prix, etc. (Tseng & Goo, 2005). Tout ceci montre l'importance de l'innovation technologique pour la création de valeur d'une entreprise (Tseng & Goo, 2005). Depuis les années 80, nous assistons à un réel changement des coûts réalisés par les entreprises pour l'innovation. Lev (2001) a mis en lumière les données de ce déplacement des investissements des entreprises des ressources tangibles vers des ressources intangibles. Dans son étude, les ressources tangibles d'une entreprise représentaient en moyenne 62% de leur valeur en 1982 et 20% à l'année 1998. Les entreprises ont compris cette importance des ressources intangibles pour la création de valeur à cette époque (Lev, 2001). D'après (Alegre

& Lapiedra, 2005 ; Subramaniam & Youndt, 2005), la capacité d'innovation d'une entreprise dépend fortement de ses ressources intangibles, de la connaissance qu'elle possède et évidemment de la façon dont elle les utilise. C'est pourquoi l'investissement en recherche et développement (R&D) est nécessaire mais pas suffisant pour augmenter les capacités d'innovation de l'entreprise. Cet investissement en R&D doit être combiné avec un investissement dans le CI de l'entreprise, notamment dans le CH pour que les employés parviennent à transformer ce potentiel d'innovation en innovations réelles, dans le CO afin de contrôler le processus des ressources possédés ainsi que dans le CS pour la connaissance acquise et transférée avec les collaborateurs (Mariz Perez et al., 2012).

2.1.2 Les modélisations récentes des processus d'innovation

Avec le changement technologique dû à la transformation numérique, le processus d'idéation et la gestion des connaissances en entreprise se transforment également. Auparavant, les modèles de processus d'innovation étaient essentiellement linéaires avec des étapes précises de la création de la technologie avec les connaissances scientifiques à la commercialisation de cette technologie sur le marché. Désormais, les processus d'innovation sont plus complexes, dits hybrides. La commercialisation d'une innovation est réussie après l'analyse de plusieurs dimensions comme l'environnement, l'utilisateur, le design, la technologie, la propriété intellectuelle ou encore le modèle d'affaires. Dans la modélisation du processus d'innovation il s'agit à la fois d'interpréter le rôle individuel des personnalités innovantes et aussi d'interpréter les différents liens, relations et partages avec les acteurs à l'échelle de l'organisation et son contexte avec le réseau d'organisations. Dans le modèle développé par Cohendet et al. (2010), les acteurs dans le processus d'innovation sont représentés par trois ensembles appelés *underground*, *middle-ground* et *upperground*. L'ensemble des individus au potentiel créatif à la base de l'innovation est appelé « *underground* », tandis que le réseau des entreprises qui forment l'écosystème d'une industrie est appelé « *upperground* ». La liaison entre ces deux ensembles s'effectue par un intermédiaire appelé « *middle-ground* » qui se matérialise par un groupe d'individus. Le *middle-ground* contribue aux effets d'externalité, à faire circuler et partager les connaissances entre les différentes entités qu'elles soient individuelles ou institutionnelles. Ce *middle-ground* nécessite une place, il peut s'agir d'un rassemblement physique ou virtuel. Ces trois acteurs réalisent différentes tâches dans le développement de nouvelles innovations. L'*underground* réalise le processus d'idéation, le *middle-*

ground contribue à la gestion des connaissances et l'*upperground* réalise la gestion des projets de ces innovations. L'*underground* qui concentre le Capital Humain dans le processus d'innovation est donc un acteur nécessaire à la concrétisation de nouvelles innovations. La Figure 2-2 illustre les relations entre les trois dimensions d'acteurs dans le milieu organisationnel.

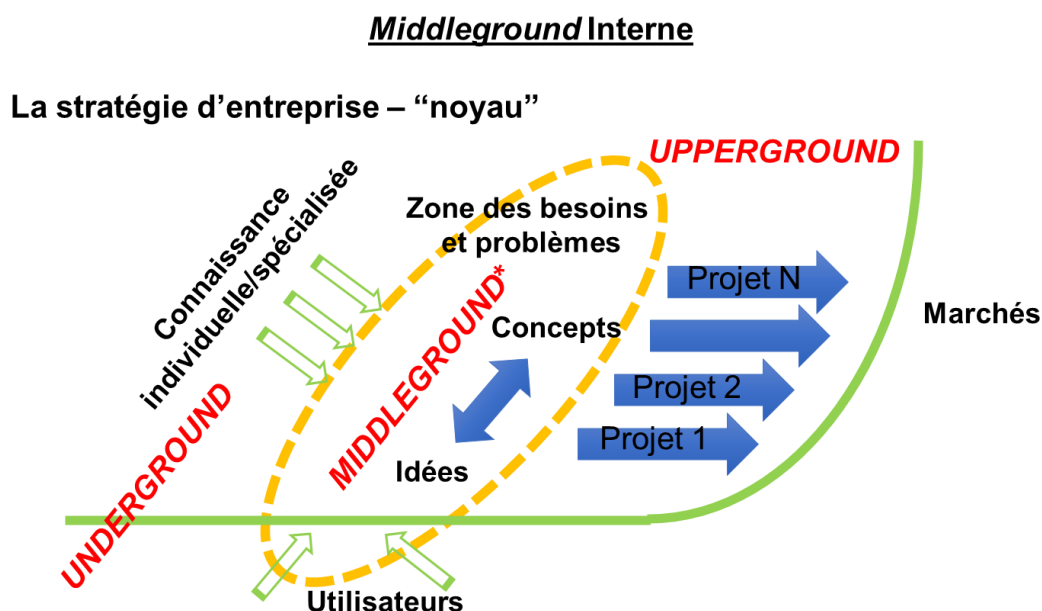


Figure 2-2 Structure interne du *Middle-ground* dans le milieu organisationnel. Adapté de (Cohendet & Simon, 2020)

A travers cette schématisation du processus d'innovation en milieu organisationnel, la frontière de couleur verte représente la délimitation à la fois physique et immatérielle de ce qui est interne à l'organisation d'un côté et de ce qui est externe à l'organisation de l'autre. Premièrement, il est intéressant de noter que du côté interne de l'entreprise, la connaissance individuelle à l'échelle de l'*underground* est une ressource majeure pour alimenter les activités de résolution de problèmes à l'échelle du *middle-ground*. Lors de cette étape de résolution de problèmes, les idées collectives et les concepts théoriques permettent l'avancement des projets de l'organisation au niveau de l'*upperground* qui deviennent des innovations dès lors que les prototypes sont finalisés pour être positionnés sur le marché. L'ensemble de ces étapes déterminent la stratégie d'entreprise qui constitue le « noyau » du processus d'innovation. Deuxièmement, cette structure démontre que chaque acteur : l'*underground*, le *middle-ground* et l'*upperground* ne se limitent nullement aux ressources internes de l'organisation. On observe en effet que les « utilisateurs » sont des parties prenantes du processus d'innovation et contribuent aux événements d'idéations qui s'opèrent à

travers le *middle-ground*. On pourrait expliquer ce rôle comme une rétroaction après utilisation du prototype afin d'apporter des informations supplémentaires et bénéfiques à l'aboutissement de l'invention. D'ailleurs, le *middle-ground* lui-même a également lieu dans des espaces tels que des forums, des webinaires ou encore dans des *laboratoires*, des incubateurs qui peuvent être des endroits physiques ou virtuels externes à l'organisation. Le *middle-ground* représenté en pointillés orange, s'étend hors de la frontière verte de l'organisation dans la Figure 2-2. Enfin, l'*upperground* comprend également le réseau d'organisations qui entourent l'entreprise qui réalise ses projets et qui interagissent en tant que collaborateurs, clients, fournisseurs, investisseurs, partenaires, etc. Pour résumer le rôle des trois acteurs dans le processus d'innovation : l'*underground* explore les idées, tandis que le *middle-ground* converti les idées et l'*upperground* exploite les idées. Cependant, cette modélisation récente du processus d'innovation est essentiellement appliquée dans les milieux culturels et artistiques jusqu'à présent. Toutefois, ce modèle semble s'appliquer au milieu industriel de façon plus générale, comme nous venons de le présenter et comme Brunet & Malas (2019) le présentent pour le secteur de la santé. L'objet de ce projet de recherche est d'approfondir la nature des ressources individuelles à l'échelle de l'*underground* qui ont un impact dans le processus d'innovation. Il semble évident que la connaissance individuelle, telle qu'elle est présentée, joue un rôle moteur dans le processus d'innovation toutefois les interactions à tous les niveaux entre les trois dimensions laissent aussi penser que d'autres compétences à l'échelle individuelle sont motrices dans ce processus d'innovation. C'est pourquoi nous nous intéressons à présent au concept de Capital Humain.

2.2 Le Capital Humain

Comme nous l'avons évoqué en introduction, le Capital Humain (CH) est une des composantes du Capital Intellectuel (CI) d'une entreprise à laquelle s'ajoutent le Capital Organisationnel (CO) et le Capital Social (CS) (Subramanian & Youndt, 2005). On retrouve également les termes de « *managerial cognition* » (Helfat & Martin, 2015) ou de « *structural capital* » (Mariz Perez et al., 2012) pour référer au concept du CO et « *relational capital* » (Mariz Perez et al., 2012) pour le CS. Toutefois, quelques soient leurs appellations dans la littérature du CI, les trois concepts : CH, CO et CS sont toujours présents. Le concept du CH, à l'initiative du processus d'innovation, est mis en valeur pour la première fois dans « *Principles of the political economy* » (1890), Alfred Marshall et toujours cité plus d'un siècle après par Levitt (1976), Alfred Marshall considérait que

« *the most valuable from all the capital is the one invested in the human being* ». Le rôle du Capital Humain dans le processus d'innovation et par conséquent sur la croissance économique d'une entreprise dans une économie basée sur la connaissance est indéniable (Becker, 1964 ; Subramanian & Youndt, 2005 ; Mariz Perez et al., 2012). Dans cette partie, nous définissons le concept du CH et quelles sont les dimensions sous-jacentes de ce concept qui interfèrent avec le potentiel d'innovation d'une entreprise. Dans le *Global Human Capital Report* (2017), le Forum Economique Mondial (FEM) définit le Capital Humain comme les connaissances et les compétences que l'homme possède pour créer de la valeur dans le système économique. Il s'agit de caractériser comment ces connaissances et ces compétences à l'échelle individuelle permettent de contribuer au processus d'innovation.

2.2.1 Le rôle du Capital Humain (CH) dans le processus d'innovation

Mariz Perez et al. (2012) effectuent une analyse descriptive des modèles sur le CI qui formalisent notamment le lien entre les dimensions du CH avec les capacités d'innovation de l'organisation. A travers ces modèles récents du CI en entreprise listés dans le tableau ci-dessous, le CH est interprété de différentes manières à la fois comme une ressource tangible et intangible dans l'organisation. De plus, selon les modèles, le CH est perçu soit comme une source d'innovation directe soit comme une part intégrante de l'organisation qui augmente la valeur économique de l'entreprise. Pour les modèles du CI qui s'inscrivent dans ce deuxième cas de figure, le CH est un facteur de l'organisation qui interagit avec la capacité d'innovation : soit le CH est une source indirecte dans le processus d'innovation, soit l'innovation augmente le CH de l'entreprise. À l'image des caractéristiques des modèles sur le CI présentés ci-dessous, les études ne sont pas unanimes sur le rôle du CH dans le processus d'innovation. À partir de la recension de littérature effectuée par Mariz Perez et al. (2012) sur les modèles du CI, les caractéristiques du processus d'innovation, les dimensions du CH et le lien entre le CH et les capacités d'innovation de l'entreprise sont précisés dans le Tableau 2-1.

Tableau 2-1. Modèles théoriques sur le CI en entreprise²

<i>Théorie du CI</i>	<i>Caractéristiques du modèle</i>	<i>Dimensions et mesures du CH</i>	<i>Lien</i>
<i>Integral Credit Scorecard Model</i> (Kaplan & Norton, 2001)	<ul style="list-style-type: none"> • L'innovation est perçue comme un processus interne qui augmente la valeur du CI (CH, CS et CO) 	<ul style="list-style-type: none"> • Le CH représente la disponibilité de compétences, de talents, de savoir-faire des employés pour établir des processus internes à l'entreprise qui garantissent le succès stratégique 	• Non
<i>Skandia Navigator Model</i> (Roos et al., 1997)	<ul style="list-style-type: none"> • La performance de l'entreprise vient de sa capacité à créer de la valeur durable selon sa vision stratégique et la mission de l'entreprise • L'innovation est directement liée à la contribution du CH de l'organisation 	<ul style="list-style-type: none"> • Le capital humain inclue les capacités individuelles, la connaissance, les compétences et l'expérience des employés et managers 	• Oui
<i>University of Western Ontario Model</i> (Bontis, 1998)	<ul style="list-style-type: none"> • L'innovation est le résultat des processus d'apprentissage et de communications individuels qui ont lieu dans l'entreprise • Le CI est la somme du CH, capital structural et capital relationnel • Le CH est source d'innovation, de rénovation stratégique et de valeur pour l'entreprise 	<ul style="list-style-type: none"> • Le CH est fait d'un stock de connaissances de types tacites et explicites 	• Oui
<i>Intangible Assets Monitor</i> (Sveiby, 1997)	<ul style="list-style-type: none"> • Le CI de l'entreprise est considéré comme une ressource entièrement intangible • L'innovation est un facteur qui dépend de ces trois composantes : croissance et renouvellement, efficacité, stabilité de chaque capital du CI • Seulement les experts participent au CH de l'entreprise 	<ul style="list-style-type: none"> • Le CH est composé de 3 dimensions • Pour la croissance / renouvellement : expérience, niveau d'éducation, coûts de formation, rotation des employés • Pour les indicateurs d'efficacité : proportion d'experts, valeur ajoutée par les experts • Indicateurs de stabilité : moyenne d'âge, niveau de service et rémunération par poste, rotation des experts 	• Non

² Adapté de Mariz-Perez et al. (2012).

Tableau 2-2. Modèles théoriques sur le CI en entreprise (suite et fin)

<i>Théorie du CI</i>	Caractéristiques du modèle	Dimensions et mesures du CH	Lien
Nova Model (Camison et al., 2000)	<ul style="list-style-type: none"> Le CI est déterminé par 4 composantes : le CH, le CO, le CS et le capital d'innovation et d'apprentissage 	<ul style="list-style-type: none"> Le CH s'explique par les connaissances individuelles (tacites et explicites) qui sont divisées en : connaissances techniques, l'expériences, les compétences de leadership, les compétences de travail d'équipe, la stabilité des employés et les compétences de management 	Oui
Intellect Model (Euroforum, 1998)	<ul style="list-style-type: none"> Le CI est divisé en 3 blocs : CH, CO et capital relationnel. Le CI est à mesurer à travers le temps et aussi comme potentiel Le CH est une partie de la valeur de l'entreprise et le capital organisationnel doit être intégré dans la mesure de la capacité d'innovation 	<ul style="list-style-type: none"> Le CH est la connaissance tacite et explicite utile pour l'organisation et que l'entreprise possède à travers ses équipes. Le CH est aussi la capacité des employés à créer de la connaissance <ul style="list-style-type: none"> Dimension actuelle : satisfaction personnelle, typologie de l'employé, compétences des employés, leadership, travail d'équipe, stabilité et risque de perte de stabilité Dimension future : l'amélioration des compétences de l'employé et la capacité d'innovation des personnes et des équipes 	Non
Stewart Model (Mount et al., 1998)	<ul style="list-style-type: none"> Le CI est la somme du CH, du capital relationnel et du capital client Le CH et le capital client sont des ressources volatiles. Les équipes interdisciplinaires apprennent, formalisent et capitalisent le savoir-faire individuel en le partageant pour le rendre moins dépendant de l'individu <p>Le CH et le capital structural se renforce mutuellement lorsqu'ils créent de la valeur pour le client et alignés avec la stratégie d'entreprise</p>	<ul style="list-style-type: none"> Le CH comprend le talent et les compétences des employés. 	Oui
Intellectus Model (Bueno et al., 2004)	<ul style="list-style-type: none"> Le CI est la somme du CH, capital structural (organisationnel et technologique) et capital relationnel (commercial et social) <p>L'innovation est générée à travers le capital technologique et l'innovation agit comme un lien entre les valeurs internes provenant du CH et CO et les valeurs externes liées au capital commercial et CS</p>	<ul style="list-style-type: none"> Le CH fait référence à la connaissance explicite ou tacite et individuelle ou social ; utilisée par l'employé ou les groupes afin de créer de la connaissance et de l'utiliser lors des missions de l'entreprise. 	Non

Le Tableau 2-1, nous montre que le CH est toujours présent dans la structure du CI. Néanmoins, le rôle du CH dans le processus d'innovation diffère selon chaque modèle. On distingue deux types d'analyses, les modèles qui intègrent le concept d'innovation soit comme résultat du processus d'innovation, soit comme capacité d'innovation. D'une part, les modèles qui intègrent l'innovation comme résultat du processus d'innovation sont les quatre modèles suivants : *Integral Credit Scorecard Model*, *University of Western Ontario Model*, *Intangible Asset Monitor*, *Intellectus Model*. D'autre part, le Tableau 2-1 contient des modèles du CI en entreprise qui intègrent la capacité d'innovation en tant que potentiel : *Skandia Navigator Model*, *Nova Model*, *Intellect Model* et *Steward Model*. Ce sont ces quatre modèles qui permettent d'analyser le lien entre le CI et la capacité d'innovation qui sont intéressants dans le cadre de ce projet. En effet pour cette étude exploratoire, il s'agit de s'intéresser plus finement au lien entre le concept du CH et le potentiel d'innovation, également appelé, capacité d'innovation du CO en entreprise. L'OCDE dans le Manuel d'Oslo (2018, p.173) recommande d'ailleurs de porter attention au lien étroit entre le CH et les capacités d'innovation des entreprises afin d'améliorer la gestion des ressources humaines, notamment de la main-d'œuvre hautement qualifiée. En somme, il s'agit d'approfondir le lien précis entre le CH et la dimension de potentiel d'innovation du CO que l'on retrouve dans le modèle du CI en entreprise *Skandia Navigator Model* établi par Roos et al. (1997). D'après cette même recension de littérature Mariz Perez et al. (2012), il résulte de la théorie du CI que le CH d'une entreprise peut être perçu au premier abord comme une ressource tangible, c'est-à-dire des employés qui possèdent un stock de connaissances. Toutefois, afin d'interpréter le rôle du CH dans le processus d'innovation de l'entreprise, ces modèles théoriques révèlent aussi l'importance des propriétés intangibles du CH, notamment les capacités des employés à améliorer leurs connaissances ou de les exploiter pour développer de nouvelles innovations.

2.2.2 Les indicateurs du Capital Humain (CH)

De la recension de littérature sur les modèles du Capital Intellectuel en entreprise, détaillés dans la section ci-dessus, il en résulte que ces modèles comportent des indicateurs de mesures du CH de façon non structurée (sans dimension) ou alors partielle. Le cadre conceptuel pour le CH a été établi par le FEM en 2017 dans le *Global Human Capital Report* (GHCRC) est probablement le modèle de la littérature le plus complet actuellement sur la modélisation du CH. C'est ce cadre conceptuel que nous utiliserons comme référence. La plupart des publications sur le concept du CH propose

des indicateurs de mesure sans distinguer les différentes dimensions du Capital Humain. Afin d'apporter davantage de précision, le FEM représente le concept du CH à l'échelle globale, pour mesurer le CH de la population de tout un pays, par quatre dimensions qui sont la capacité, le développement, le déploiement et le savoir-faire. Ce cadre conceptuel est d'ailleurs réutilisé par Abuzyarova et al. (2019) pour analyser le lien entre le CH d'une part et les sciences, la technologie et l'innovation d'autre part. Dans le *GHCR* du FEM (2017), quatre dimensions du CH sont définies la capacité, le déploiement, le développement et le savoir-faire.

Dimension 1 : la capacité

L'objectif de la dimension de capacité du capital est d'évaluer la quantité de connaissance que possède la population. Selon le FEM, une population plus éduquée est mieux préparée à s'adapter aux nouvelles technologies, innover et être compétitif à un niveau global dans le modèle capitaliste. En effet, la complexité de nombreuses innovations modernes nécessite la mise en commun et l'intégration de plusieurs volets des connaissances individuelles (Subramanian & Youndt, 2005). Dans le *GHCR*, la dimension de capacité du CH est quantifiée par quatre mesures communes du niveau d'éducation formelle, ventilée à travers les groupes d'âges dans la population active. Ces indicateurs mesurent le pourcentage de la population qui a acquis au moins le niveau d'éducation primaire, le niveau d'éducation secondaire et le niveau d'études supérieures respectivement et la proportion de la population qui a un niveau basique en littératie et numératie.

Dimension 2 : le déploiement

La deuxième dimension du CH développée dans le *GHCR* est celle du déploiement. Dans le milieu du travail, le CH est valorisé par l'apprentissage par la pratique, la connaissance tacite et les échanges entre collègues et l'apprentissage formel sur le tas. Le déploiement permet d'évaluer aussi bien l'aptitude des personnes à participer activement dans leur milieu de travail que la participation de certains segments de la population pour des secteurs de métiers spécifiques. Les indicateurs du déploiement du CH sont le taux de participation de la population active, l'écart des genres en matière d'emploi, le taux de chômage et le taux de sous-emploi. Le taux de participation la population active est la mesure la plus large de l'étendue de participation de la population sur le marché du travail : la mesure inclue la population en recherche d'emploi et la population qui travaille, la mesure sur l'écart des genres en matière d'emploi est incluse car le sujet reste une faiblesse majeure dans de nombreux secteurs notamment pour les emplois qualifiés à travers le

monde comme le mentionne le FEM dans le *Global Gender Gap Report* FEM(2017), le taux de chômage est la part de la population active sans emploi et le taux de sous-emploi est la part de la population actuellement employée qui pourrait et souhaiterait travailler davantage.

Dimension 3 : le développement

La dimension du développement du Capital Humain concerne l'éducation des générations futures de la population active et l'amélioration et la reconversion de la génération de travailleurs actuelle. Le développement du CH est composé de neuf indicateurs dans le *GHCR*. Cette dimension est mesurée par l'accès à l'éducation pour les enfants et les jeunes, la génération future de travailleurs, capturée par le taux net d'inscription à l'école primaire et le taux net d'inscriptions à l'école secondaire, ainsi que le taux global d'inscription aux études supérieures (*tertiary studies*). Les jeunes adultes qui ont validé leurs études secondaires, ont le choix de poursuivre vers des études supérieures, afin d'acquérir des compétences spécialisées pour leur vocation ou d'entrer directement sur le marché du travail. Cette dimension du déploiement comporte aussi deux indicateurs quantitatifs sur la qualité de l'éducation primaire et la réponse aux besoins du marché du travail de l'éducation primaire, mesurée par la communauté exécutive de chaque pays. De plus, la dimension du développement comporte une analyse fine des compétences acquises par la population, capturée par une évaluation de la diversité des compétences des jeunes diplômés pour visualiser l'éventail d'expertises disponibles par pays. Un éventail de compétences varié et riche permet au pays d'assurer sa résilience et son adaptation face au changement exponentiel technologique et économique en cours (Brynjolfsson & McAfee, 2014). Le dernier indicateur est celui de la formation tout au long de la vie (*lifelong learning*) des employés dans les entreprises, qui permet d'apercevoir le renouvellement et l'amélioration des compétences des employés.

Dimension 4 : le savoir-faire

La quatrième dimension du concept de CH est le savoir-faire, il concerne l'étendue et la profondeur des compétences spécialisées utilisées pour chaque métier. Le savoir-faire comporte les indicateurs de complexité économique, le taux d'emploi de la population hautement et moyennement qualifiée, ainsi que la disponibilité des employés qualifiés, qui représente la facilité ou la difficulté à pourvoir les postes dans les entreprises. L'indicateur de complexité économique, quant à lui, se mesure par le degré de sophistication de la « connaissance productive » d'un pays qui peut être observée empiriquement par la qualité de son exportation de produits.

Le cadre conceptuel sur le CH développé par le FEM couvre 21 indicateurs de mesures uniques qui permettent d'évaluer les différentes dimensions du Capital Humain selon cinq tranches d'âge : 0 à 14 ans, 15 à 24 ans ; 25 à 54 ans ; 55 à 64 ans et 65 ans et plus. Dans le cadre qui nous intéresse, à savoir l'**impact du CH en entreprise** sur le processus d'innovation, il est important de repérer parmi ces quatre dimensions du CH et plus particulièrement parmi ces indicateurs, lesquels sont appropriés à la population cible en entreprise. Ainsi, on pourra considérer que ce sont les trois tranches d'âges de 15 à 24 ans, 25 à 54 ans et celle de 55 à 64 ans, au plus large, qui auront un intérêt pour l'étude du CH en entreprise. Dans le tableau ci-dessous, qui figure dans le *GHCR*, on retrouve dans la colonne de gauche les dimensions et les indicateurs et dans l'en-tête les tranches d'âges qui permettent de ventiler la mesure des indicateurs selon les activités des groupes de population. Le Tableau 2-3 permet de repérer aisément en couleur bleue les indicateurs de mesure du CH pour les tranches d'âge d'intérêt.

D'après le Tableau 2-3, cette modélisation du concept du CH permet d'analyser de manière globale par pays à l'échelle mondiale sur l'ensemble de la population aux différentes tranches d'âge ainsi que pour l'ensemble des secteurs d'activités. Pour l'étude exploratoire de ce mémoire, nous retiendrons les dimensions du CH et par conséquent les indicateurs qui présentent un intérêt pour l'analyse du lien entre le CH des employés et potentiel d'innovation du CO de l'entreprise dans le cadre d'un modèle à l'échelle d'une entreprise qui œuvre dans un secteur industriel. Ainsi les indicateurs des dimensions de la capacité, du déploiement et du savoir-faire du CH peuvent être mesurés pour une population d'employés en entreprise, tandis que la dimension de développement du CH s'applique partiellement pour une population d'employés en entreprise avec les indicateurs de taux d'inscription dans des programmes de formations professionnalisantes, taux d'inscription en études supérieures, diversité des compétences des diplômés, la qualité du système d'éducation et l'étendue de la formation tout au long de la vie des employés (*lifelong learning*).

De plus, afin de compléter ce cadre conceptuel du CH pour le cas d'une population d'employés en entreprise d'un secteur industriel, le modèle sur le contenu des professions O*NET (2018) permet de renforcer la mesure des dimensions de développement et de savoir-faire. D'une part, le modèle O*NET apporte un indicateur sur l'expérience professionnelle pour la dimension du développement pour les tranches d'âge de 15-24 ans, 25-54 ans et 55-64 ans. D'autre part, le modèle O*NET permet de préciser les indicateurs sur les compétences ciblées selon les professions. Les compétences des employés qui peuvent jouer un rôle sur le potentiel d'innovation

du CO d'une entreprise sont détaillées dans la sous-section 2.3.2. Afin de mener notre étude exploratoire, il semble nécessaire d'identifier quels indicateurs pourront être mesurés à partir des données LinkedIn.

Tableau 2-3. Structure du concept du CH

Dimensions	Indicateurs	Tranches d'âge				
		0-14	15-24	25-54	55-64	65+
Capacité	Littéracie et numératie		X	X	X	X
	Taux d'accomplissement d'éducation primaire		X	X	X	X
	Taux d'accomplissement d'éducation secondaire		X	X	X	X
	Taux d'accomplissement d'études supérieures			X	X	X
Déploiement	Taux de participation de la population active		X	X	X	X
	Écart des genres en matière d'emploi		X	X	X	X
	Taux de chômage		X	X	X	X
	Taux de sous-emploi		X	X	X	X
Développement	Taux d'inscription à l'éducation primaire	X				
	Qualité des écoles primaires	X				
	Taux d'inscription à l'éducation secondaire	X				
	Écart des genres en éducation secondaire	X				
	Taux d'inscription dans des programmes de formations professionnalisantes		X			
	Taux d'inscription en études supérieures		X			
	Diversité des compétences des diplômés		X			
	Qualité du système d'éducation		X			
	Étendue de la formation des employés			X	X	
Savoir-faire	Proportion des emplois à haute qualification			X	X	
	Proportion des emplois moyennement qualifiés			X	X	
	Complexité économique			X	X	
	Disponibilité des employés qualifiés			X	X	

Source : *The Global Human Capital Report. Preparing People for the Future of Work*, 2017 ©2017 par The World Economic Forum.

2.3 Les sources de données du Capital Humain en entreprise

Les études qui étudient le lien entre le CH des employés et l'innovation en entreprise sont nombreuses (Bozbura, 2004 ; Gamerschlag, 2013 ; Massingham et al., 2011). En ce sens, les modélisations du CI en entreprise recensées par Mariz Perez et al. (2012) et détaillées dans la section 2.2.2 suggèrent de consolider les outils de mesure des attributs du concept du CH. Le concept du CH semble donc difficile à mesurer dans son ensemble, notamment dû au caractère intangible de certaines dimensions du CH (Claver-Cortés et al., 2015). Il convient alors de déterminer les sources de données pertinentes et légitimes pour obtenir les informations pour mesurer le CH en entreprise. De plus, les outils utilisés pour la gestion des ressources humaines en entreprise, généralement constitués d'organigrammes et d'outils internes de *reporting*, présentent leurs limites lorsqu'il s'agit d'analyser les expertises et les connaissances de leurs employés (Wang et al., 2013). Afin d'améliorer leurs systèmes d'informations, les entreprises sont de plus en plus enclines à utiliser les médias sociaux pour leurs activités (Pisano et al., 2017). En particulier l'utilisation du réseau social professionnel de LinkedIn comme base de données du CH abrite un grand potentiel d'exploitation (Georgescu et al., 2018). En quoi le réseau professionnel LinkedIn pourrait-il améliorer les activités de ressources humaines en entreprise ? Est-il possible de constituer une structure de données du CH exploitable à partir des données LinkedIn pour évaluer le potentiel d'innovation du CO en entreprise ? Jusqu'à présent nous retrouvons quatre études qui proposent des usages des données du CH des employés de LinkedIn dans le but d'améliorer certains procédés de gestion de la connaissance en entreprise (Reese et al., 2018 ; Georgescu et al., 2018 ; Wang et al., 2013 ; Pisano et al., 2017). Georgescu et al. (2018) proposent un usage des données de LinkedIn afin d'améliorer les processus de recrutement. Wang et al. (2013) élaborent un modèle d'évaluation de l'expertise des employés à partir d'une analyse sémantique des données de compétences des profils sur LinkedIn. Pisano et al. (2017) analysent l'impact du CH sur les performances financières de l'entreprise par un modèle de régression linéaire. Enfin Reese et al. (2018) analysent l'impact du partage des types de compétences au sein de l'équipe fondatrice sur les performances de la nouvelle entreprise. Les sous-sections suivantes portent sur le réseau professionnel LinkedIn et sur les compétences du CH à l'ère du numérique. Nous y présentons l'intérêt d'utiliser les données LinkedIn pour apporter des informations complémentaires aux outils actuels de gestion des ressources humaines en entreprise, tels que les Progiciels de Gestion Intégrés, en vue d'identifier le potentiel d'innovation du CO l'entreprise à partir des données du CH.

2.3.1 Le réseau professionnel LinkedIn

LinkedIn est le réseau social (*Social Network Site*) le plus utilisé dans le monde par les professionnels (Wang et al., 2013). LinkedIn compte environ 756 millions d'utilisateurs³ en 2021. Depuis les dernières années, le réseau social LinkedIn est également devenu le centre d'attraction des entreprises. Les entreprises utilisent largement LinkedIn pour transformer leurs performances d'affaires et de production (Pisano et al., 2017). Les quatre grandes catégories d'utilisateurs du réseau social LinkedIn sont les suivantes : les membres individuels qui cherchent un nouvel emploi ou qui cherchent à se connecter à leur réseau de relations professionnelles, les spécialistes de ressources humaines qui s'appuient sur LinkedIn pour trouver les bons candidats pour leurs offres d'emploi, les spécialistes de marketing qui souhaitent atteindre une audience particulière et filtrée par des paramètres professionnels et les collaborateurs d'affaires en entreprise ou les chercheurs qui souhaitent avoir accès à une base de données professionnelle (Georgescu et al., 2018). La plus grande utilisation de LinkedIn du côté des entreprises est notamment celle pour le recrutement. Néanmoins, la catégorie d'utilisation de LinkedIn comme base de données professionnelle présente un vivier d'informations très riche pour améliorer les processus de création de valeur des entreprises (Pisano et al., 2017). En particulier Pisano et al. (2017) a souligné que les entreprises étaient de plus en plus intéressées par le contenu des données sur le CH présent dans LinkedIn. Comme le réseau professionnel possède une quantité de données très élevée sur les profils des employés à travers le monde entier, LinkedIn se positionne comme source de données du CH en entreprise à haut potentiel. Par cette étude exploratoire, il s'agit de déterminer la qualité des données du CH que l'on trouve sur le réseau LinkedIn pour identifier le potentiel d'innovation du CO d'une entreprise d'un secteur industriel. Wang et al. (2013), p.1, valorisent l'utilisation des réseaux sociaux professionnels tels que LinkedIn dans les entreprises multinationales afin d'améliorer le management des ressources humaines :

“In particular, social networking naturally captures information about the activities, interactions, and knowledge of employees in a digital form that can be mined for insight

³LinkedIn Pressroom - LinkedIn Corporate Communications - About Us - 27 avril 2021 - FY21 Q3 report Microsoft
©LinkedIn Corporation 2021

and business process improvements (Ehrlich et al., 2007; Chelmiss et al., 2012; Lin et al., 2012; Varshney et al., 2013).

Knowledgeable employees are not interchangeable because they each have specialized expertise and skills; this has an important role in a corporation's business, especially in service industries with much client interaction. Hence, it is critical to capture and understand the individual specialties of employees for successful human capital management and operation in large enterprises. For instance, if the expertise of each employee can be comprehensively catalogued, the assignment of projects can be extremely efficient and accurate since the desired experts can be called upon to meet clients' needs."

Ce passage de la publication de Wang et al. (2013) cible avec précision les enjeux de notre étude exploratoire : les projets d'innovation en R&D des entreprises multinationales d'un secteur industriel demandent des profils d'expertise très particuliers. Si l'entreprise dispose de méthodes supplémentaires lui permettant de constituer les équipes les plus adaptées aux objectifs de leurs projets d'innovation à partir des données du CH de leurs employés, alors les performances⁴ de l'entreprise s'amélioreront. Les données présentes sur LinkedIn qui capturent les activités, les interactions entre les membres du réseau et les compétences à partir des profils individuels, sont propices à l'application de méthodes d'Extraction de Connaissances à partir de Données (ECD). Les données personnelles sur les expériences professionnelles et associatives, les projets, les formations, les compétences, les recommandations professionnelles entre collègues qui sont présentes sur LinkedIn sous forme de sections sur chaque profil des membres, comportent une forme qui permet l'application de méthodes d'ECD telles que la fouille de texte, les règles d'associations, les méthodes d'analyses statistiques ou encore les méthodes d'apprentissage.

2.3.2 Les compétences du Capital Humain à l'ère du numérique

La notion de compétences du personnel figure comme un facteur clé des activités d'innovation d'une entreprise selon l'OCDE (2018). Dans le *Manuel d'Oslo* de 2018, les pratiques de gestion

⁴ Le terme « performances » est utilisé ici en tant que concept multidimensionnel. Pour une entreprise, on parle de performance économique (compétitivité), financière (rentabilité), juridique (solvabilité), organisationnelle (efficience) ou encore sociale. (Issor, 2017)

des ressources humaines comprennent la formation et le développement des compétences individuelles qui ont un impact potentiellement favorable sur les activités d'innovation de l'entreprise. De plus, la notion de compétences est toujours présente dans la définition du CH, on le retrouve dans chacune des modélisations du CI détaillées dans la section 2.2.1. Il en est de même pour le cadre conceptuel du CH développé le FEM dans le *GHCR* (2017) détaillé dans la section 2.2.2 où le CH fait référence aux connaissances et compétences que les personnes possèdent qui leur permettent de créer de la valeur dans le système et que ces personnes sont capables d'améliorer au cours du temps. Quelles sont alors les compétences individuelles recherchées pour les activités d'innovation d'une entreprise multinationale en pleine transformation numérique ? Pour l'étude exploratoire de ce projet, on s'intéresse particulièrement aux compétences individuelles recherchées pour une entreprise de l'industrie des services professionnels, scientifiques et techniques d'après la classification des secteurs industriels du code SCIAN. Ce contexte est important dans la mesure où une entreprise de ce secteur intégrera certains types de professions. Dans la suite de cette étude, il convient alors de s'appuyer sur la Classification Internationale Type des Professions (CITP-08) fournie par l'Organisation internationale du travail (OIT) car chaque emploi est défini « par un ensemble d'attributs liés aux tâches, aux activités, aux connaissances requises, aux compétences technologiques et générales, ainsi qu'aux aptitudes et valeurs personnelles » (OCDE, 2018, p.129). L'OCDE suggère notamment de se référer au modèle de contenu des professions O*NET (2018) pour dresser l'état des lieux des compétences individuelles présentes dans l'entreprise pour but d'identifier les capacités d'innovation de l'entreprise. Le modèle O*NET distingue trois groupes d'attributs qui impactent les activités d'innovation de l'entreprise :

1. Les attributs permanents qui influent sur l'**efficacité**, tels que :
 - Les capacités cognitives, en particulier l'**aptitude à produire des idées** et à raisonner ;
 - La souplesse et la **capacité d'adaptation au changement** ;
2. Les aptitudes qui facilitent la réalisation d'activités propres à certains métiers, telles que :
 - Les **compétences sociales**, qui permettent d'œuvrer avec d'autres à la réalisation d'objectifs ;
 - Les capacités de **résolution de problèmes complexes**, qui permettent de résoudre des problèmes inédits et aux contours mal définis dans un cadre concret et complexe ;

- Les **compétences techniques** qui permettent de concevoir, de mettre en place, de faire fonctionner des machines ou systèmes technologiques et d'en corriger les dysfonctionnements ;
- Les **compétences systémiques**, qui permettent de comprendre, de suivre et d'améliorer les systèmes sociotechniques ;

3. Les **valeurs professionnelles et styles de travail**, qui concernent notamment l'esprit d'entrepreneuriat, le travail d'équipe, la créativité et l'autonomie.

Les études qui énumèrent et justifient le développement de certaines compétences au niveau individuel favorables aux activités d'innovation à l'ère du numérique avec les Technologies de l'Information et de la Communication (TIC) au XXI^e siècle sont nombreuses (Subramaniam & Youndt, 2005 ; O*NET, 2018 ; Abuzyarova et al., 2019 ; van Laar et al., 2017 ; Reese et al., 2017 ; Georgescu et al., 2018 ; Claver-Cortès et al., 2015 ; Gamerschlag, 2013). Pour notre étude exploratoire qui porte sur les entreprises qui œuvre en STI, les experts de l'OCDE (2011) ciblent six groupes de compétences individuelles qui sont particulièrement adéquates pour l'identification des capacités d'innovation : les **compétences techniques** (ex : adoption des TIC), les **qualités de recherche**, les **connaissances spécifiques au secteur**, les **compétences générales** (ex : esprit critique) et les **compétences non techniques** (ex : communication, travail d'équipe). En parallèle, le réseau professionnel LinkedIn propose sa propre classification des compétences individuelles en fonction des données déclarée par les membres. Comme notre étude exploratoire porte sur l'évaluation du potentiel de LinkedIn comme source de données du CH, il est intéressant de croiser les groupes de compétences formés automatiquement par le réseau social LinkedIn avec les groupes de compétences retrouvés dans la littérature. L'enjeu d'évaluer le potentiel de LinkedIn comme source du CH est double, il s'agit d'apporter des éléments de réponses aux deux questions suivantes : La structure des données développée par le réseau social LinkedIn est-elle pertinente pour apporter des connaissances sur le potentiel d'innovation d'une entreprise multinationale de services professionnels, scientifiques et techniques ? Le contenu des données du CH disponible sur LinkedIn par les utilisateurs apporte-t-il des connaissances pour identifier le potentiel du CO d'une entreprise ? Sur LinkedIn, un membre qui complète son profil peut déclarer jusqu'à 50 compétences spécifiques et au-delà de 3 compétences inscrites, le réseau professionnel LinkedIn classe automatiquement les compétences par catégorie : « principales compétences », « connaissances du secteur », « outils et

technologie », « compétences interpersonnelles », « autres compétences »⁵. Lors de cette étude, il s'agit donc d'explorer différentes structures de groupes de compétences afin d'éventuellement identifier le potentiel d'innovation de l'entreprise par l'application de différentes méthodes d'ECD, à savoir des règles d'association et des méthodes de classification. Le chapitre suivant introduit le cadre de la recherche avec l'entreprise industrielle choisie pour le cas d'étude, les objectifs de recherche ainsi que les modèles étudiés.

⁵LinkedIn Corporation (03 août 2021). *Ordre d'affichage des recommandations de compétences*. LinkedIn. <https://www.linkedin.com/help/linkedin/answer/35681/ordre-d-affichage-des-recommandations-de-competences?lang=fr>

CHAPITRE 3 CADRE DE LA RECHERCHE

3.1 Contexte et question de recherche

Cette section présente en deux sous-sections le contexte et la question de recherche spécifiques à ce projet ainsi que les objectifs établis pour répondre à la question de recherche.

3.1.1 Contexte et question de recherche du projet

Ce projet explore un nouveau **mécanisme de soutien à l'innovation** pour des entreprises industrielles par le développement d'un modèle d'apprentissage permettant d'identifier le potentiel d'innovation du CO de l'entreprise à partir du CH des employés. Ce projet réalise une étude exploratoire pour l'entreprise CAE. L'entreprise CAE est un chef de file mondial sur le marché des simulateurs de vols et développe depuis quelques années des produits et services dans le domaine de la santé. Ces deux écosystèmes, de la santé et de l'aérospatiale, connaissent de profondes transformations numériques et sont parmi les écosystèmes d'innovation étudiés dans la littérature. Cette transformation affecte plusieurs des activités de CAE, notamment l'amélioration des infrastructures TI, l'acquisition de données en temps réel et l'analyse de données pour les machines industrielles ou encore pour l'automatisation de procédés. A propos du mécanisme de soutien à l'innovation, CAE souhaite **optimiser la constitution de ses équipes** affectées à répondre à certaines problématiques identifiées par l'entreprise, ce qui lui permettrait d'être encore plus innovante. Ainsi, la question de recherche est d'identifier quelles caractéristiques des profils sont les plus appropriées pour répondre, par exemple, à des questions techniques ou de gestion ? Par ailleurs, l'entreprise s'interroge sur les capacités transversales de ses équipes, qui sont actuellement divisées entre les départements de la santé et de l'aérospatiale. En somme, l'objectif est de pouvoir comprendre les caractéristiques des profils les plus innovants.

Pour répondre à de telles questions, CAE souhaite explorer les différentes possibilités qu'offrent maintenant l'analyse de données massives. La méthodologie envisagée, détaillée dans le chapitre 4, combinerait la fouille et l'analyse de texte dans le but de concevoir des indicateurs qui permettraient à l'entreprise de faire des suivis en temps réel autant que possible. Ce travail sortira, dans un deuxième temps, des murs de l'entreprise pour explorer les collaborations de CAE avec d'autres acteurs des écosystèmes de l'aérospatiale et de la santé. Cette deuxième approche adoptera une combinaison de méthodologie, dont un sondage et des entrevues.

Comme présenté dans le paragraphe 2.1.2, les acteurs de l'innovation dans une industrie peuvent être identifiés et analysés à différentes échelles. En partant de l'acteur individuel, des communautés ou des groupes : *underground*; en passant par les réseaux, les collectifs : *middle-ground*; et allant jusqu'aux firmes innovatrices, réseaux de firmes, clusters et organisations culturelles : *upperground*. Ces catégories d'acteurs présentent des propriétés caractéristiques et relatives à leur échelle qui permettent d'examiner le processus d'innovation (Simon, 2009). Les qualités humaines des individus influent directement sur leur créativité et leur contribution au processus d'innovation. Il est intéressant de mettre en perspective la modélisation du processus d'innovation de Simon (2009) avec l'écosystème d'innovation de l'industrie aéronautique dont fait partie l'entreprise CAE.

En combinant l'utilisation des paramètres de ces études à des indicateurs du Capital Humain de l'ordre personnel, il semble alors possible de concevoir un modèle avec des paramètres qui offrent une analyse à l'échelle des employés pour la conversion d'idée en connaissance organisationnelle, autrement dit de la conversion d'idées des employés en des ressources qui présentent une valeur économique ajoutée pour l'entreprise. Plus précisément, il s'agit d'analyser la relation entre l'*underground* et le *middle-ground* par des indicateurs du Capital Humain des employés, acteurs de l'*underground*, et des indicateurs mesurant l'efficacité du processus de génération et de conversion d'idée en ressource organisationnelle nécessaire pour l'innovation. À titre comparatif ce genre de modélisation a déjà été réalisé dans le cadre d'optimiser le personnel impliqué pour le cas du système de santé (Brunet & Malas, 2019).

3.1.2 Objectifs

L'objectif général du mémoire est d'établir un mécanisme de soutien à l'innovation avec le développement d'un modèle d'apprentissage pour identifier le potentiel d'innovation du Capital Organisationnel (CO) d'une entreprise d'un secteur industriel à partir du Capital Humain (CH) des employés.

Les sous-objectifs de cette étude exploratoires sont les suivants :

- SO1. Obtenir les mesures des indicateurs du CH et du potentiel d'innovation du CO adaptés au contexte de l'entreprise CAE à partir des sources de données secondaires ;

- SO2. Établir une structure standard des données collectées du CH et du potentiel d'innovation du CO exploitable pour des modèles d'apprentissage ;
- SO3. Déterminer les méthodes d'apprentissage efficaces pour caractériser le potentiel d'innovation pour l'ensemble d'entraînement (ou d'apprentissage) utilisé des données des employés de l'entreprise CAE, avec respect de la confidentialité des données personnelles ;
- SO4. Évaluer le potentiel de LinkedIn comme source de données du CH pour la ou les méthodes d'apprentissage retenues.

3.2 Contribution originale

Ce projet souhaite apporter des connaissances nouvelles sur le concept du CH et son lien direct avec le concept d'innovation de l'entreprise. Premièrement, ce projet permet de contribuer à la modélisation du processus d'innovation dans un milieu organisationnel élaboré par Cohendet et al. (2010) dans le cas d'une entreprise d'un secteur industriel, avec l'étude exploratoire pour l'entreprise CAE qui œuvre dans le secteur aérospatial et le secteur de la santé. Dans la littérature de Cohendet et al. (2010), la modélisation du processus d'innovation avec les groupes de l'*underground*, le *middle-ground* et l'*upperground*, est définie initialement dans le cadre de la ville créative, des milieux culturels et artistiques. Les organisations telles que l'entreprise CAE ou leurs employés pourraient ne pas se projeter en tant que « communauté artistique ou culturelle » qui trouve son inspiration dans la « vie de nuit ». Pourtant cette modélisation du processus de l'innovation de Cohendet et al. (2010) semble pouvoir expliquer les processus d'innovation qui ont lieu au sein d'une entreprise industrielle telle que CAE et de son écosystème. D'ailleurs, cette modélisation du processus d'innovation par l'*underground*, le *middle-ground* et l'*upperground* a déjà été appliquée une première fois au secteur de la santé dans *L'innovation en santé : Réfléchir, agir et valoriser* (2019). Lors de ce projet nous contribuons plus précisément à la compréhension de l'interaction entre l'*underground* et le *middle-ground*, notamment pour le processus de conversion des idées. Le modèle que nous développons dans ce projet permettrait alors d'identifier les liens entre les indicateurs du CH disponible dans l'*underground* et le potentiel d'innovation présent dans le *middle-ground*, c'est-à-dire le potentiel d'innovation présent dans le CO de l'entreprise. Deuxièmement, le caractère innovant du projet réside dans la nature du concept d'innovation analysé. Lors de ce projet nous souhaitons concevoir et explorer la validité d'un

modèle qui, à partir d'indicateurs du CH, permette d'évaluer le potentiel d'innovation du CO de l'entreprise. Dans la littérature sur les modèles qui analysent le lien entre le CI d'une entreprise et l'innovation, le concept d'innovation est mesuré de manière globale soit comme « résultats d'innovation » ou bien comme « capacité d'innovation » (Mariz Perez et al., 2012). Nous proposons de caractériser plus finement le potentiel d'innovation dans le cadre conceptuel de cette étude, du fait que les différents types d'innovation requièrent différents types de connaissances et de compétences (Cardinal, 2001). Le modèle a pour vocation de fournir un mécanisme de soutien à l'innovation à l'entreprise pour disposer d'une meilleure compréhension de l'impact des indicateurs du CH des employés sur le potentiel d'innovation du CO de l'entreprise. Le projet a donc pour finalité d'optimiser les activités d'innovation et favoriser la croissance de l'entreprise. Troisièmement, le modèle développé lors de ce projet permet d'apporter une nouvelle source de données du CH des employés, à savoir le réseau social professionnel LinkedIn. Wang et al. (2013) ont identifié le réseau professionnel LinkedIn comme source de données pour des méthodes d'ECD. Ces méthodes fondées sur l'analyse de données viennent en soutien aux outils de gestion de ressources humaines pour améliorer certains procédés telles que la constitution d'équipes pour améliorer les performances des entreprises multinationales. L'étude exploratoire présente un nouvel usage des données du CH présent sur LinkedIn pour identifier le potentiel d'innovation du CO de l'entreprise. De plus, la méthode réalisée intègre une structure reproductible des indicateurs du CH construits à partir des données de LinkedIn pour d'autres entreprises de différents secteurs. Ainsi, ce projet réalisé en exploration avec l'entreprise CAE de l'industrie aérospatiale et de la santé, propose un modèle adaptable pour toute autre entreprise qui œuvre dans une autre industrie et a priori de toutes tailles car l'étude est effectuée à partir du CH individuel.

3.3 Cadre conceptuel et modèles

Cette section est réservée à préciser le cadre conceptuel de l'étude exploratoire ainsi qu'à présenter les différents modèles étudiés pour atteindre l'objectif de recherche qui est d'identifier le potentiel d'innovation du CO d'une entreprise d'un secteur industriel à partir du CH de ses employés. Il s'agit donc de cibler les dimensions des concepts de potentiel d'innovation du CO ainsi que le concept du CH des employés, afin de chercher d'éventuels liens de causalité ou liens prédictifs entre les concepts à partir de données sur un ensemble d'employés de l'entreprise.

D'une part, de la revue de littérature, nous retenons que le potentiel d'innovation du CO d'une entreprise industriel est notamment marquée par son activité de publications de brevets. En effet, pour rappel, **le CO est la connaissance institutionnelle et l'expérience codifiée qui est utilisée et réside dans les bases de données, les manuels, les structures, les systèmes, les procédés ainsi que dans les brevets** (Youndt et al., 2004).

D'autre part, la revue de littérature nous a permis d'analyser le rôle du CH de l'entreprise pour ses activités d'innovation à la section 2.2.1. Dans le cadre de cette étude exploratoire, nous avons retenu qu'il s'agit d'approfondir le lien précis entre le CH et la dimension de potentiel d'innovation du CO que l'on retrouve dans le modèle du CI en entreprise *Skandia Navigator Model* établi par Roos et al. (1997).

Dans cette optique, nous avons analysé le concept du CH dans la littérature pour en retenir les dimensions qui sont susceptibles de présenter des liens avec le potentiel d'innovation du CO de l'entreprise industrielle. D'après l'étude du modèle du FEM dans le *Global Human Capital Report* (2017), les dimensions de Capacité, de Déploiement et de Savoir-faire peuvent se mesurer dans une population d'employés en entreprise. De plus, le modèle O*NET approfondit la dimension de savoir-faire définissant des catégories de compétences spécifiques aux professions dans le modèle de contenu des professions O*NET (2018). Nous avons détaillé les compétences utiles à la dimension du savoir-faire du CH dans la section 2.3.2.

Compte tenu de cette analyse de la littérature nous étudierons trois modèles afin d'identifier le potentiel d'innovation du CO en entreprise à partir du CH de ses employés.

Les schémas ci-dessous présentent les modèles qui seront testés lors de l'étude exploratoire, ces modèles décrivent le cadre conceptuel et les variables utilisées pour mesurer les dimensions de chaque concept. La méthodologie utilisée pour mener l'étude exploratoire, le choix des variables le traitement des données, la création de variables puis les méthodes d'analyses de données seront développées dans le Chapitre 4.

La

Figure 3-1 représente le 1^{er} Modèle à variables binaires pour l'ensemble des dimensions du CH (savoir-faire et développement) et du potentiel d'innovation du CO afin d'appliquer la méthode des règles d'association décrite à la section 4.5.1. La Figure 3-2 montre le 2^{ème} Modèle à variables binaires et discrètes sur lequel sera appliqué l'arbre de décision, décrit à la section 4.5.2. Enfin la

Figure 3-3 représente le modèle à variable binaires, discrètes et catégoriques qui permettra l'application de l'algorithme des K-Moyennes pour la classification, décrit à la section 4.5.3.

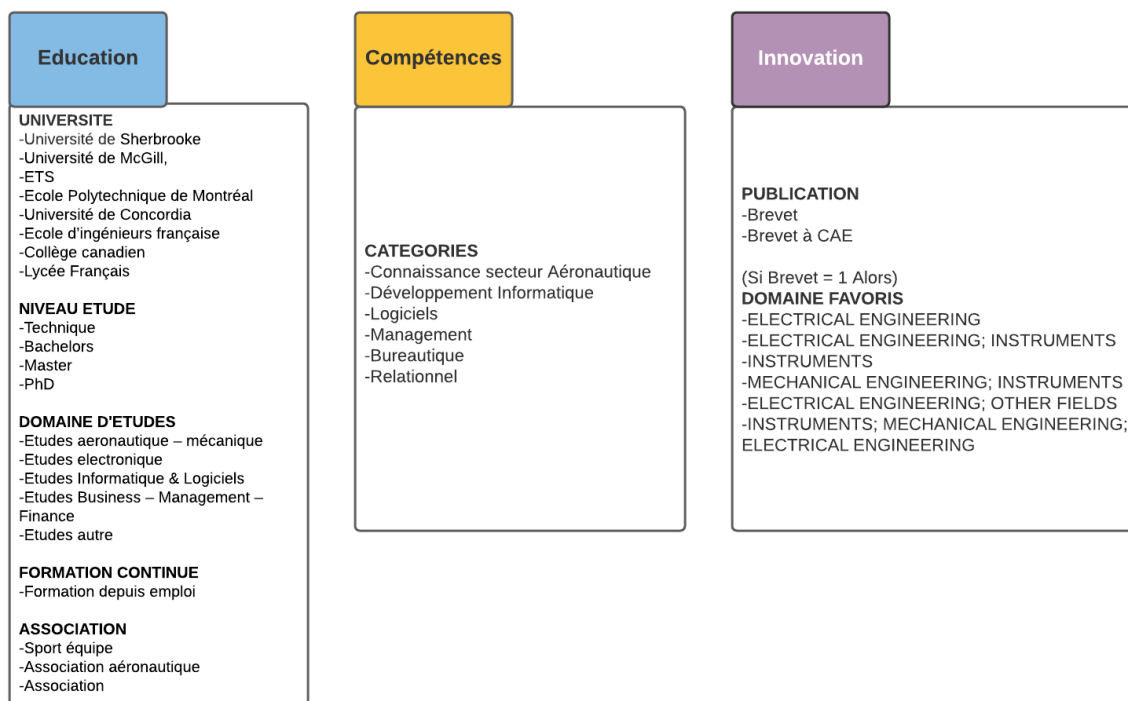


Figure 3-1. Modèle 1 de variables binaires

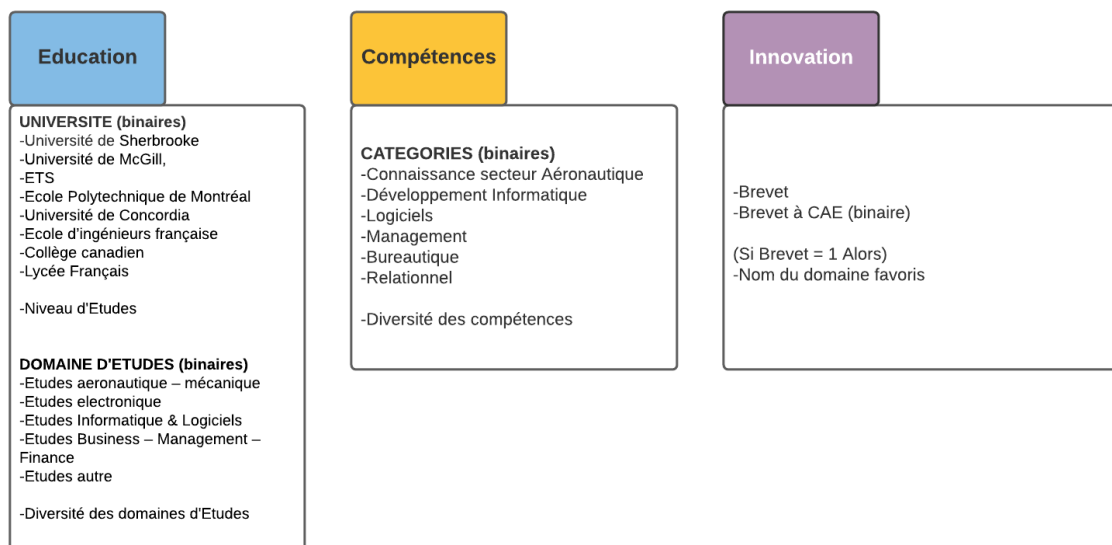


Figure 3-2. Modèle 2 de variables binaires et discrètes

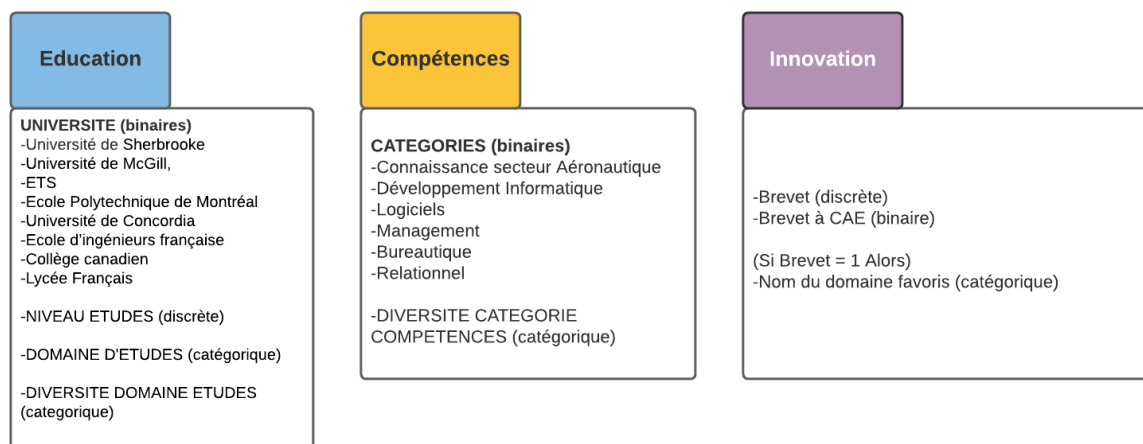


Figure 3-3. Modèle 3 de variables binaires, discrètes et catégoriques

Ces 3 modèles sont fondés sur le même cadre conceptuel, à savoir les dimensions du Développement, du Déploiement, du Savoir-faire pour le concept du CH des employés mais n'incluent pas la Complexité Économique ni la Disponibilité des employés qualifiés, d'après le FEM (2018) et le modèle O*NET (2018) et des Activités de publications pour le concept de potentiel d'innovation du CO de l'entreprise (Youndt et al., 2004). Toutefois, le 3ème modèle utilise des variables différentes (variables discrètes « Niveau d'études », variable catégorique « Domaine d'études », variable catégorique « Diversité des domaines d'études », variable catégoriques « Compétences », variable catégorique « Diversité des catégories de compétences »), pour mesurer les dimensions du CH des employés et les dimensions du CO de l'entreprise. Le premier modèle est composé uniquement de variables binaires, tandis que le second modèle comporte des variables binaires et discrètes et le troisième modèle est construit avec des variables binaires, discrètes et catégoriques. Les choix de construction de ces modèles ont fait l'objet de expérimentations qui sont présentées au Chapitre 5 et ces choix proviennent notamment de l'application de différentes méthodes d'analyse de données : des méthodes de règles d'association applicables au modèle 1, des arbres de décisions applicables au modèle 2 et des méthodes de classification type algorithme K-Moyennes dans le modèle 3. Ces méthodes d'analyse de données sont décrites à la section 4.5.

CHAPITRE 4 MÉTHODOLOGIE

Ce chapitre est réservé à la description détaillée de la méthodologie adoptée lors de cette étude exploratoire. Lors de ce chapitre, il s'agit de présenter l'ensemble des étapes de l'étude exploratoire afin de pouvoir reproduire les analyses effectuées et de pouvoir enrichir les méthodes développées au cours de ce projet. Tout d'abord la première section sur la collecte de données décrit la manière dont les données ont été obtenues à partir la base de données de brevets Patseer et à partir du réseau professionnel LinkedIn pour les données du CH des employés. La deuxième section établit la méthode de sélection élaborée pour l'étude exploratoire, qui décrit précisément la population pour laquelle ont été collecté les données. La troisième section présente les variables sélectionnées parmi l'ensemble des attributs disponibles de part et d'autre des sources de données (Patseer et LinkedIn). La quatrième section décrit les méthodes de traitement de données de la source de données Patseer ainsi que de la source de données LinkedIn obtenue par Moissonnage afin de fournir les outils nécessaires à la reproduction de l'étude exploratoire. La cinquième section détaille la création des différentes variables utilisées dans les différents modèles de cette étude exploratoire. La sixième section décrit les méthodes d'analyse de données établies lors de l'étude exploratoire qui sont les règles d'associations, les arbres de décisions et l'algorithme de classification des K-Moyennes.

4.1 Données

Cette étude utilise deux sources de données pour collecter premièrement les données brevets réalisés par des employés actuels dans l'entreprise ou des employés ayant quitté l'entreprise à partir de la base Patseer et deuxièmement les données du CH des employés de l'entreprise sur le réseau professionnel LinkedIn d'autre part. La plateforme Patseer ainsi que le réseau professionnel LinkedIn constituent des sources de données secondaires. Nous verrons dans une première sous-section l'intérêt et la valeur ajoutée de s'orienter vers le logiciel Patseer pour les données brevets, notamment le nombre d'attributs et d'information à explorer. Ensuite dans une deuxième sous-section, nous présenterons les attributs du CH qui peuvent être collectés à partir des profils des membres LinkedIn.

4.1.1 Extraction d'information des données brevet

Afin de collecter les données brevets réalisés par l'entreprise CAE, nous optons pour le logiciel Patseer⁶. Patseer est une « plateforme intelligente » développée avec des algorithmes d'Apprentissage Machine et d'Intelligence Artificielle pour collecter et ajouter de la valeur aux données brevets du monde entier. La plateforme Patseer intègre les données de plus de 136 millions de brevets à travers le monde, de plus de 108 autorités, 96 millions d'archives de texte intégral des brevets. La plateforme fournit notamment de la valeur ajoutée aux données brevets avec les revendications indépendantes, les résumés et objets d'inventions, les descriptions d'état de l'art, les descriptions des dessins pour chaque brevet ainsi qu'une cartographie des brevets – domaines technologiques et une cartographie brevets – secteurs industriels. À partir des recherches de brevets qu'il est possible d'effectuer sur Patseer par inventeurs ou entreprises, le logiciel fournit plus de 100 attributs de données par brevet. La plateforme Patseer propose d'extraire les attributs de données à la convenance de l'utilisateur parmi cette centaine d'attributs. Parmi la centaine d'attributs proposés par Patseer, nous retenons 94 attributs pour l'extraction qui sont détaillés dans le tableau ci-dessous. Ces attributs de données qui possèdent différents types : nombres, caractères et temporels, permettent un traitement de données afin de construire les variables exploitables pour les méthodes de classification et les règles d'associations envisagée lors de l'étude exploratoire.

⁶ <https://patseer.com/>

Tableau 4-1. Liste des attributs des données brevets collectés sur Patseer

Nom du groupe d'attributs	Nom du sous-groupe d'attributs	Noms des attributs
Brevets – Domaine Technologique	Catégories	<ul style="list-style-type: none"> Industry, Tech Domain, Tech Sub Domain, Field of search, ICR, CPC, US Class (UC), ICO, Locarno Class, FI, FTERM
	Identification	<ul style="list-style-type: none"> Record Number, Application No., Application No. Original, PCT Application Number, PCT Patent Number Title, Abstract, Summary of Invention, Description, Description of Drawings, Advantages of Invention
	Structure Technologique	<ul style="list-style-type: none"> Backward Citation Count, Backward Citations, Backward Citations with Details, Forward Citations, Forward Citations with Details, No. of Forward Citations (Individual), No. of Forward Citations (All Simple Family Members), No. of Forward Citing Families, No. of non-self-citations (forward citation), No. of non-self-citations (backward citation), No. of References, References
Légal	Dates- Brevets	<ul style="list-style-type: none"> Legal Status Date, Filing/Application Date, Publ Date, Last Modified Date Estimated Expiry Date, Earliest Priority Date, Earliest Publication Date, Priority Details, Priority Date
	Dates- Structure Technologique	<ul style="list-style-type: none"> Simple Family Members (Publication Date), Simple Family Publication Dates, Simple Family ID, SFAM All countries, Family Legal Status (Dead/Alive), Extended Family Members (Publication Date), Extended Family Publication Dates, Extended Family ID, No. of Simple Family Members, No. of Extended Family Members
	Statut Légal 2020	<ul style="list-style-type: none"> Record Type, Legal Status (Dead/Alive), Legal Status Current, INPADOC Legal Status, Register Legal Status, Register
	Géographie	<ul style="list-style-type: none"> Publication Country, Designated States, Active in Designated States Priority Country Code, Priority Year
	Inventeurs	<ul style="list-style-type: none"> Inventors, First Named Inventor, Inventor Address, Inventor Country, No. of Inventors
Relations	Cessionnaire	<ul style="list-style-type: none"> Assignee, Assignee Address, Current Assignee, Current Owner, No. of Assignees, Assignee Country, Assignee Original, Assignee State, Assignee City, Applicant
	Examineur	<ul style="list-style-type: none"> Examiner, Attorney, Attorney Address

À partir des catégories d'attributs provenant de l'extraction de données brevets de la plateforme Patseer, il est possible de créer un nombre important de variables concernant les brevets d'une entreprise. Voici une liste d'idées d'utilisation des données et/ou méthodes d'ECD applicables pour ces données brevets du tableau ci-dessus provenant de l'outil Patseer :

1. Le regroupement des brevets par les attributs de domaines technologiques et sous-domaines technologiques établis par Patseer à partir du code IPC (*International Patent Classification*) et détaillés dans la section 4.3.
2. La collection des données du CH des employés ayant réalisé des brevets et l'identification des entreprises du même écosystème avec lesquelles former des partenariats intéressants à partir des données textuelles sur les inventeurs, les cessionnaires, les titulaires des brevets.
3. Des méthodes de traitement du langage naturel (ou *Natural Language Processing* en anglais – NLP) sur les données textuelles du titre, du résumé, du résumé de l'invention, de la description, des avantages de l'invention afin de déceler des « catégories » de brevets potentiellement différentes des classifications de domaines technologiques IPC.
4. La création d'indicateurs de brevets à partir des références : références précédentes, références suivantes, familles de brevets associées à chaque brevet afin de positionner la technologie du brevet analysé et déterminer son influence sur le marché.
5. L'évaluation potentielle de l'importance de la technologie accordée par l'entreprise par le financement de son brevet à partir des attributs de données de type temporel permettant d'obtenir de l'information sur les dates de dépôts des brevets, leurs durées de validités légales.

Lors de cette étude exploratoire, nous approfondirons les idées 1 et 2 qui sont le regroupement des brevets par domaines et sous-domaines technologiques en fonction des codes IPC ainsi que la collection des données du CH à partir des inventeurs de brevets de l'entreprise industrielle étudiée. Les étapes de construction des indicateurs et de recherche des profils des employés de l'entreprise provenant de ces deux cas d'usage seront détaillés dans les sections suivantes.

Dans un premier temps, il s'agit de déterminer les variables des données brevets obtenus à partir de la plateforme Patseer. Comme nous l'avons présenté, la plateforme Patseer permet l'extraction de brevets d'une entreprise, en particulier pour l'entreprise CAE qui nous intéresse pour cette étude exploratoire. Notre extraction de données brevets au format .csv nous permet d'obtenir un tableau de 94 attributs par brevet, directement exploitable pour les expérimentations futures. Toutefois, l'objectif de la recherche est d'identifier le potentiel d'innovation du CO de l'entreprise à partir du CH de ses employés. Il s'agit donc d'utiliser des variables brevets qui décrivent spécifiquement les activités du CO de l'entreprise. Dans le cadre de l'étude exploratoire, il s'agit d'utiliser les attributs du domaine technologique des brevets afin de cibler les activités réalisées par les employés de l'entreprise CAE. **Pour cela, nous optons pour le choix des attributs sur le Domaine Technologique et Légal, conformément aux catégories d'attributs décrits dans le Tableau 4-1.** Nous utilisons les attributs des sous-catégories d'attributs : Catégorie et Inventeurs. Les attributs de la sous-catégorie « Catégorie » comprennent notamment le domaine technologique (*Tech Domain*), le sous-domaine technologique (*Tech Sub Domain*), le champ de recherche (*Field of search*), le code IPC (*International Patent Classification*⁷). Les attributs de domaine technologique et de sous-domaine technologique des brevets sont construits à partir des code IPC, leur construction est détaillée dans la sous-section 4.4.1. Les attributs de la sous-catégorie « Inventeurs » comprennent notamment le nom des inventeurs qui permet à la fois de chercher le profil LinkedIn et d'en extraire les données du CH si l'accord de la personne est donné sur le réseau professionnel LinkedIn et aussi de dénombrer le nombre de brevets réalisés par la personne et d'obtenir de façon transposée par un traitement de données, les données de chaque brevet réalisé par la personne.

En pratique, l'extraction au format .csv des données provenant de Patseer fournit une table de données comportant des variables brevet pour chaque brevet. Or la table de données souhaitée après le traitement de données brevets doit fournir la liste des brevets et les Catégories de ces brevets réalisés pour chaque inventeur. En effet, il est nécessaire que les données brevets soient compatibles avec les données du CH pour chaque employé de CAE Inc., c'est-à-dire que dans la

⁷ La classification internationale des brevets se trouve sur le site Internet suivant : [Lien vers la classification WIPO](#)

table de données à exploiter nous ayons les indicateurs du CH et les indicateurs des données brevets en fonction des inventeurs.

Pour ce faire, dans un premier temps à partir de l'extraction au format .csv des données brevets de Patseer il s'agit d'identifier les inventeurs de brevets qui ont un profil LinkedIn et qui acceptent l'utilisation des données de leurs profils pour l'étude exploratoire. Puis il s'agit d'abord de transformer la table des données provenant de Patseer vers une table de données qui donne les brevets avec leurs attributs réalisés par chaque inventeur de l'entreprise CAE Inc.

4.1.2 Extraction d'information des données LinkedIn

Dans la littérature, le réseau professionnel LinkedIn est identifié comme une source de données du CH à haut potentiel pour améliorer les procédés des ressources humaines en entreprise (Wang et al., 2013). En effet, LinkedIn est un réseau où les utilisateurs ont la possibilité de publier un vivier riche d'informations professionnelles. Il s'agit de déterminer les variables des données du CH des employés de l'entreprise que nous obtenons sur les profils LinkedIn, avec pour rappel le respect de la confidentialité des données personnelles de ces personnes. Sur le réseau professionnel LinkedIn, les profils des employés comportent la même structure. **Les profils LinkedIn, qui s'apparentent un Curriculum Vitae digitalisé, comportent des sections nommées Introduction, Infos, Sélection de contenus, Parcours, Compétences, Réalisations et Informations supplémentaires.** De façon plus détaillée, la rubrique « Introduction » laisse un champ de texte libre et optionnel à l'utilisateur pour se présenter. La rubrique « Infos » constitue l'en-tête des profils LinkedIn, elle est systématique présente sur tout profil et résume le poste actuel de la personne, le dernier établissement académique fréquenté par l'utilisateur ainsi que d'autres informations tels que l'emplacement géographique de la personne et son nombre de relations sur le réseau professionnel. La rubrique « Sélection » de contenus présente les interactions, les publications de l'utilisateur sur le réseau professionnel. Ces informations ne font pas l'objet de notre étude exploratoire. La rubrique « Parcours » inclue les expériences professionnelles, la formation regroupant l'établissement académique, le niveau d'études, le domaine d'études et les éventuelles associations durant le parcours, et les licences et certifications ainsi que les expériences de bénévolat. Cette rubrique peut être renseignée librement par l'utilisateur mais constitue néanmoins les données sur lesquels nous effectuons la sélection, expliqué à la section 4.1.2. Ainsi, nous savons qu'en utilisant les filtres de l'outil de recherche de LinkedIn, les profils qui résultent

de la requête auront des champs complétés dans cette rubrique « Parcours ». Cette information est non négligeable car elle indique que les données de la rubrique « Parcours » pour notre ensemble ne comporte pas de données manquantes, c'est-à-dire que chaque profil de l'ensemble présentera au moins une valeur dans cette rubrique « Parcours ». Les profils LinkedIn peuvent également comporter une rubrique « Compétences » qui compote toute son importance dans la mesure du concept du CH vu dans la revue de littérature. Le réseau LinkedIn propose sa propre catégorisation de l'attribut des compétences, donc on verra par la suite la pertinence. Ensuite, la rubrique « Réalisations » permet de renseigner les publications, les brevets, les cours, les projets, les prix et distinctions, les résultats d'examens, les langues parlées, les organisations et les causes soutenues par l'utilisateur. Enfin la rubrique « Informations supplémentaires » permet d'afficher sur son profil les recommandations professionnelles reçues par d'autres utilisateurs.

Ainsi, pour les attributs du CH de cette étude exploratoire, nous décidons de retenir les variables relatives aux « Infos » par rapport à l'emplacement de la personne, son sexe et sa catégorie d'âge; les variables relatives aux « Parcours » sur la formation et les expériences professionnelles; ainsi que les variables sur les « Compétences ».

4.1.3 Résumé de la sélection des informations

Le tableau suivant permet de résumer le choix des variables effectué, qui ont été jugé pertinents en s'appuyant sur la littérature pour la construction d'un modèle qui permette de répondre à l'objectif d'identifier le potentiel d'innovation du CO à partir du CH des employés de l'entreprise qui œuvre dans un secteur industriel.

Tableau 4-2. Résumé de la sélection des variables.

Source	Variables	Description
Brevet	Domaine Technologique	Cet instrument de mesure permet de désigner les domaines et sous-domaines technologiques ainsi que le code IPC afin de spécifier les domaines techniques des inventions.
	Légal	Cet instrument de mesure permet de désigner les inventeurs de brevet et d'assurer leur conformité selon les réglementations de brevet.
LinkedIn	Infos	Cet instrument de mesure permet d'obtenir l'emplacement, la tranche d'âge et le genre des employés avec leur accord; et de faire lien avec les données brevets des inventeurs.
	Parcours	L'instrument de mesure de Parcours permet de retracer les expériences professionnelles et la formation de l'employé.
	Compétences	L'instrument de mesure de Compétences permet de mesurer les champs de compétences aussi bien techniques (<i>hardskills</i>) que sociales (<i>softskills</i>) des employés de l'entreprise. Un instrument de mesure sur la dimension des compétences est nécessaire pour caractériser le concept du CH.

D'une part les variables des données brevets permettent de cibler les inventeurs, leur domaine technologique d'invention et par extrapolation leur tendance à produire des brevets dans un seul domaine ou dans une pluralité de domaines. D'autre part, les variables des données de LinkedIn permettent la mesure du CH par la série de variables offerte sur le réseau professionnel.

4.2 Sélection des données

Les données de cette étude ont été sélectionnées par choix raisonné. Le choix porte sur la sélection des employés de l'entreprise CAE dont on souhaite collecter les données du CH et des données brevets si ces employés sont inventeurs de brevets. Dans cette section, il s'agit de préciser les critères de recherche des profils des employés afin de pouvoir reproduire la table de donnée à l'état brut, sans traitement de données à ce stade, pour l'entreprise CAE ou pour reproduire l'étude exploratoire pour toute autre entreprise d'un secteur industriel des services professionnelles, scientifiques et techniques. Afin de comprendre le choix de l'ensemble de données, il s'agit en premier lieu de comprendre les étapes des méthodes d'apprentissage envisagées par la suite. L'étude exploratoire teste des **méthodes de classification semi-supervisée** dont l'objectif est de trouver des groupes d'appartenance, appelés classes, non connue à l'avance à chacune des « observations » qui représentent les employés dans le cadre de cette étude. Ainsi les méthodes de classification semi-supervisée utilisent un ensemble de données afin de leur attribuer une classe d'appartenance, dont l'objectif de notre étude est d'identifier le potentiel d'innovation pour

chacune des classes selon des descripteurs. Les descripteurs que nous utilisons pour identifier le potentiel d'innovation du Capital Organisationnel de l'entreprise sont les données brevets provenant de la plateforme Patseer. Dans le cadre de l'étude, les descripteurs du potentiel d'innovation sont les indicateurs pour chaque classe déterminée par l'algorithme du **nombre de brevets moyens réalisés par les employés** et le **nombre de brevets moyens dans le domaine technologique favoris de l'inventeur**. Ces deux indicateurs permettront d'attribuer un potentiel d'innovation à chaque classe d'appartenance, non connue à l'avance, déterminée par le modèle de classification. Dans un premier temps, afin de pouvoir mettre en œuvre un modèle de classification semi-supervisée il s'agit d'obtenir un ensemble de profils employés de l'entreprise CAE qui permettent également d'obtenir une mesure des descripteurs pour le potentiel d'innovation. Cet ensemble composé des employés qui sont inventeurs de brevets constituera l'**ensemble d'entraînement**. Ceci signifie que l'ensemble d'entraînement doit comporter des profils d'employés ayant déposé au moins un brevet. Dans un second temps, il s'agira d'utiliser ce modèle de classification pour effectuer de la prédiction du potentiel d'innovation à partir d'un nouvel ensemble, nommé **ensemble de test**, composé de profils d'employés de l'entreprise CAE qui n'ont pas « encore » réalisé de brevets afin de leur attribuer un potentiel d'innovation, selon leur classe prédite par le modèle de classification. En somme, la sélection des données s'effectue en deux étapes : la première consiste à former l'ensemble d'entraînement à partir des inventeurs de brevets issus de la recherche sur la plateforme Patseer croisé avec les profils du réseau professionnel LinkedIn afin de garantir la complétion des données du CH pour ces profils; puis la seconde étape qui consiste à obtenir un ensemble de test, composé de profils d'employés de l'entreprise CAE qui ont complété leur profil sur le réseau professionnel LinkedIn.

4.2.1 Sélection des données d'entraînement

L'ensemble d'entraînement des employés de l'entreprise CAE inventeurs de brevets, tel qu'il a été décrit dans l'introduction de la section 4.2, est issu de la recherche des brevets de l'entreprise CAE. Les attributs de notre base de données utilisée pour les méthodes de classification contiennent les indicateurs du potentiel d'innovation provenant des données brevets de Patseer ainsi que des données du CH issues des profils LinkedIn complétés inventeurs de brevets de l'entreprise CAE sur me réseau professionnel. Les différentes étapes pour construire cet ensemble d'entraînement sont illustrées avec la figure ci-dessous.

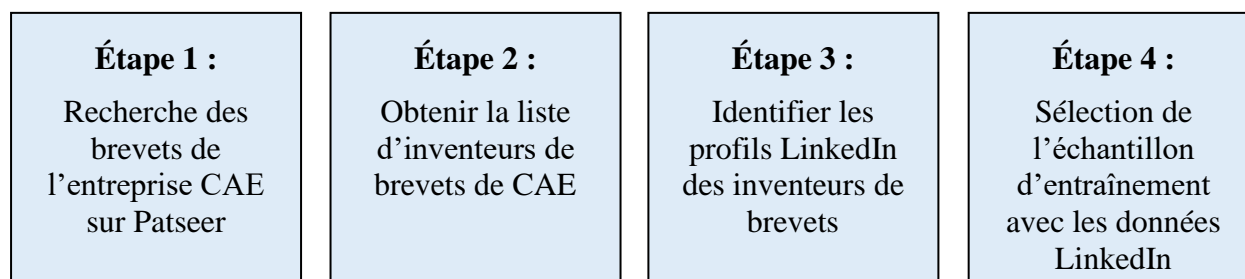


Figure 4-1. Étapes de sélection de l'ensemble d'entraînement

La première étape consiste à appliquer une requête sur la plateforme Patseer afin de collecter les données brevets déposés par l'entreprise CAE. Il s'agit tout d'abord d'identifier l'ensemble des appellations de l'entreprise CAE qui sont présents dans la base de données Patseer afin de cibler les brevets dont une entité de l'entreprise CAE est désignée comme cessionnaire. En effet, l'entreprise CAE est un acteur pluridisciplinaire et international, l'entreprise possède des entités qui portent des noms distincts et des localisations à travers le monde. Avec ce premier critère de recherche des brevets, il s'agit donc de couvrir la totalité des entités internes à l'entreprise CAE mais aussi d'exclure les organisations portant l'acronyme CAE dans leur label et qui ne sont pas des entités de l'acteur de l'aérospatial et de la santé qui concerne notre étude. Ainsi, notre requête sur la plateforme Patseer se limite à une recherche des brevets avec comme cessionnaires les entités « CAE INC », « CAE ELECTRONIC LTD », « CAE HEALTHCARE CANADA INC », « CAE HEALTHCARE INC (US) ». De plus, comme l'entreprise CAE est multinationale il s'agit de rechercher les brevets dans le monde entier : les **brevets octroyés** et les **brevets déposés** auprès de l'Office Européen des Brevets (OEB), de l'Office de la Propriété Intellectuelle du Canada (OPIC), du *United States of Patent and Trademark Office (USPTO)* et du *World Intellectual Property Organization (WIPO)*. En effectuant la recherche de brevets avec les critères précédents sur la plateforme Patseer, on trouve **408 brevets**. Les dates de dépôt des brevets vont de 1982 à l'année 2020, année à laquelle a été effectuée la collecte des données brevets sur Patseer. Ensuite, à la deuxième étape il s'agit d'obtenir la liste des inventeurs brevets parmi ces 408 brevets. On identifie **152 inventeurs différents de brevets** qui travaillent ou ont travaillé pour l'entreprise CAE. Parmi ces 152 inventeurs de brevets chez CAE, **94 personnes sont membres du réseau professionnel LinkedIn**. De plus, parmi ces 94 personnes membres du réseau LinkedIn **91 personnes ont des profils publics sur LinkedIn**, ces personnes tolèrent la consultation et l'utilisation de leur données partagées avec le réseau professionnel. Enfin, dans le cadre de cette étude exploratoire, nous

retenons uniquement les personnes employées à l'interne de l'entreprise CAE et dont CAE figure parmi leur 5 derniers postes occupés dans leur carrière professionnelle. **L'ensemble d'entraînement est finalement composé de 85 employés de l'entreprise CAE.** Notre ensemble d'entraînement a donc la particularité de posséder à la fois les données du CH provenant de LinkedIn ainsi que les données brevets provenant de Patseer, car ces 85 employés actuellement ou anciennement dans l'entreprise CAE sont également inventeurs de brevets.

4.2.2 Sélection des données de test

Dans un deuxième temps, l'ensemble de données est complété par des employés actifs de l'entreprise CAE qui possèdent un profil LinkedIn public. Il est nécessaire de posséder un compte sur LinkedIn afin d'effectuer une recherche de profils individuel sur le réseau professionnel. Lorsque l'on utilise l'outil de recherche du réseau professionnel LinkedIn pour trouver la liste des employés de l'entreprise CAE, le réseau LinkedIn nous indique environ **9 200 résultats de profils d'employés actifs de l'entreprise CAE.** A ce stade une contrainte s'impose sur le nombre de résultats qu'il est possible de consulter avec un usage gratuit du réseau LinkedIn, sans abonnement spécifique. **Les résultats issus de la requête sur le réseau LinkedIn sont limités à 1 000 profils** et apparaissent dans un ordre spécifique à l'utilisateur qui effectue la recherche : les membres qui font partis du cercle de « connexions » de l'utilisateur apparaissent en premier et s'en suivent les profils individuels qui apparaissent dans un ordre aléatoire. Il convient alors de maximiser le nombre de profils individuels obtenus avec l'outil de recherche sur LinkedIn pour obtenir l'ensemble de test le plus large possible dans l'entreprise CAE. La limite des 1000 résultats de profils individuels peut être élargie grâce à l'utilisation de filtres lors de la recherche, notamment **le filtre sur l'Université d'études des employés.** Pour les entreprises multinationales, il est courant que des groupes d'employés soient issus d'une même université ou d'une même grande école. C'est en effet le cas pour l'entreprise CAE, où selon chaque pays où est implanté l'entreprise, nous pouvons retrouver cinq universités ou grande écoles majoritaires dans les formations des employés. Dans le Tableau 4-3 nous retrouvons le nombre de résultats sur LinkedIn des employés actifs de l'entreprise CAE selon les filtres par localisation croisés avec les filtres par universités :

Tableau 4-3. Résultats de la recherche des profils sur LinkedIn⁸

							Total de profils dans les 5 universités majoritaires	Total de profils par pays
Canada	Université	Concordia University	Polytechnique Montréal	McGill University	École de technologie supérieure	Université de Sherbrooke		~4 000
	Nombre de profils	589	375	338	307	157	1 766 (>1000)	
États-Unis	Université	Embry-Riddle Aeronautical University	University of South Florida	Community College of the Air Force	University of Phoenix	University of Central Florida		~1 800
	Nombre de profils	169	160	99	46	40	514 (<1 000)	
Royaume-Uni	Université	Buckinghamshire New University	The Open University	University of Brighton	London Metropolitan University	CAE (Formation)		560
	Nombre de profils	30	16	13	7	18	84 (<1 000)	

⁸ La requête a été réalisée sur LinkedIn.com le 19 mai 2021.

Les statistiques descriptives ci-dessus nous indiquent qu'il est possible d'obtenir potentiellement plus de profils LinkedIn de l'entreprise CAE au Canada en utilisant les filtres sur les universités de formation qu'en utilisant simplement le filtre de localisation par pays. En effet, le filtre de localisation au Canada fournirait environ 4000 résultats et par conséquent il serait seulement possible d'obtenir 1000 profils distincts, tandis qu'en effectuant les requêtes pour les cinq universités de formations majoritaires pour les employés de CAE, on obtiendrait au maximum 1766 profils distincts. Cependant, il est possible que des employés aient étudié dans plusieurs de ces universités ou grandes écoles au cours de leur éducation. Ces personnes sont alors retrouvées à plusieurs reprises dans les résultats des requêtes de profils LinkedIn, ce qui réduit le nombre de profils LinkedIn d'employés distincts de l'entreprise CAE. D'autre part, une deuxième contrainte **sur la confidentialité des profils amène à réduire le nombre maximum d'individus pour l'ensemble de test** qu'il est possible d'obtenir avec l'outil de recherche. Pour l'ensemble de test nous avons utilisé l'outil de recherche avec les filtres des employés actuel de CAE, de la localisation au Canada et avec le filtre de l'École Polytechnique Montréal; parmi les 375 profils trouvés, 271 profils LinkedIn étaient publics. De plus, notre ensemble de test est également formé des résultats de la requête avec les filtres des employés actuels de CAE, de la localisation au Canada et de l'Université de Sherbrooke; parmi les 157 profils trouvés, 151 profils LinkedIn étaient publics. Lors du regroupement de ces deux ensembles de profils, nous obtenons 408 profils LinkedIn distincts d'employés actifs chez CAE. Il y a donc 14 profils qui appartiennent aux deux ensembles, car nous retrouvons seulement 408 profils distincts parmi les 422 profils potentiels ($271 + 151 = 422$). Il est possible de réitérer l'opération avec l'outil de recherche sur les autres universités ou écoles supérieures, toutefois nous décidons de conserver seulement ces 408 profils d'employés de CAE pour l'ensemble de test étant donné que notre ensemble d'entraînement n'est composé seulement de 85 profils. **L'ensemble de test est donc composé de 408 profils distincts d'employés actuels de CAE ayant effectué leurs études au minimum dans l'un des deux établissements d'enseignement supérieur : Polytechnique Montréal ou l'Université de Sherbrooke.**

4.3 Traitement des données

Cette section porte sur la description du traitement des données pour l'étude exploratoire. Elle permet de détailler le traitement des données du CH de LinkedIn. L'objectif du traitement des données est d'obtenir une table de données brevets, obtenue précédemment, compatible avec la table de données du CH obtenue avec la source de données LinkedIn. En respectant la méthode de sélection décrite précédemment à la section 4.2, il s'agit en premier lieu d'obtenir l'ensemble d'entraînement qui est composé des inventeurs de brevet de l'entreprise CAE Inc. Ainsi, étant donné que ce sont les seuls employés de l'entreprise qui sont inventeurs de brevets, l'étape de traitement des données brevet est à effectuer uniquement sur ce premier ensemble, que nous avons nommé ensemble d'entraînement. La collecte et le traitement de données pour l'ensemble de test porte sur le traitement de données LinkedIn.

- *Traitement des données LinkedIn*

L'obtention des données LinkedIn s'effectue par une méthode dite de WebScraping, qui est un processus de récupération automatique de données d'un site Web en l'occurrence LinkedIn. Cette méthode est appliquée sur les profils des employés de CAE Inc. qui possèdent un profil public et qui laissent alors libre de visibilité les données de leur profil. De plus, les données utilisées sont ensuite anonymisées de sorte que les résultats obtenus lors des expérimentations ne soient pas associés aux personnes mais uniquement interprétée par des relations de causalité et des méthodes de classification en fonction du jeu de données du CH et des données brevets.

Le principe de l'outil de moissonnage développé pour notre étude exploratoire est le suivant : le Moissonneur utilise les adresses URL qui correspondent aux résultats des requêtes de l'outil de recherche LinkedIn qui mènent aux profils identifiés lors des méthodes de sélection des données d'entraînement (4.2.1) et de test (4.2.2). Pour ensuite charger l'intégralité du code HTML du profil LinkedIn de chaque personne puis traiter les données LinkedIn du CH présentes dans le code HTML et les structurer dans un format approprié .csv. Pour ce traitement de données, nous développons 3 Moissonneurs qui sont différents et ont un principe de fonctionnement distincts. En particulier, nous avons un Moissonneur qui permet d'extraire la liste d'URL des profils LinkedIn de l'ensemble d'entraînement, un deuxième Moissonneur qui permet d'extraire la liste d'URL des profils LinkedIn de l'ensemble de test et enfin un Moissonneur plus complexe qui utilise

l'ensemble de ces adresses URL des profils LinkedIn pour en extraire les données du CH de chaque employé et les traiter dans un format structuré. Ce sont ces étapes de traitement de données que nous allons détailler dans cette partie.

Dans le cas de **l'ensemble d'entraînement**, nous possédons les noms des inventeurs de brevets en entrée du Moissonneur. Il s'agit donc d'automatiser l'outil de sorte qu'il recherche les bons profils LinkedIn des inventeurs de brevet qui sont dans l'entreprise CAE Inc. Ce Moissonneur produit à la fin de cette étape une table de données comportant la liste URL des inventeurs de brevets.

Dans le cas de **l'ensemble de test**, nous cherchons la liste d'employés à partir de l'outil de recherche de LinkedIn selon les filtres définis à la section 4.2.2. Les requêtes avec l'outil de recherche fournissent des pages qui contiennent chacune une liste de 10 employés de CAE Inc. qui répondent aux critères de recherche définis par les filtres de l'outil de recherche. A ce stade, le Moissonneur permet de balayer chaque page de résultats en extrayant les URL des profils LinkedIn publics. Ainsi, à cette étape nous obtenons une table de données qui correspond à chaque URL des profils LinkedIn des employés de CAE Inc. qui correspondent à notre requête.

Une fois ces deux premières étapes terminées, nous obtenons une liste d'URL des profils LinkedIn qui constituent notre ensemble. Il s'agit alors de passer **à l'étape de traitement des données LinkedIn** à proprement parlé. Chaque adresse URL menant vers le profil LinkedIn des employés de CAE Inc. faisant parti de notre ensemble comporte un code HTML dont la structure est propre au navigateur internet et sa version. Le développement du Moissonneur est donc maintenable à condition de s'assurer que l'outil soit à jour avec le navigateur et la structure du code HTML. Après extraction des données dans les rubriques du profil LinkedIn qui nous intéressent sur les sections Infos, Parcours et Compétences, il s'agit de structurer dans un format .csv les données du CH pour chaque employé de CAE Inc. de notre ensemble. **Ainsi, pour chaque employé le traitement de données s'effectue selon 50 variables brutes** : 2 variables pour la rubrique Info sur le genre et la tranche d'âge; 18 variables pour la formation décomposée sur les 3 diplômes les plus récents avec 6 variables identiques sur le nom de l'établissement, le niveau de diplôme, le sujet d'étude, la période, l'association et la description; 30 variables pour les expériences professionnelles décomposée sur les 5 expériences professionnelles les plus récentes avec 6 variables identiques sur le titre du poste occupé, le nom de l'entreprise, l'emplacement, les dates d'occupation du poste, la durée occupé à ce poste et une description; enfin le jeu de données brutes contient 2 variables sur

les Compétences qui sont les meilleures compétences (*Top Skills*) une liste des 3 meilleures compétences de l'employés selon les notation de son réseau professionnel et les autres compétences (*Other Skills*) qui est une liste de compétences divisée en catégories préétablie par LinkedIn à savoir *Industry Knowledge*, *Tools & Technologies*, *Interpersonal Skills*, *Other Skills*. Les jeux de données brutes obtenus, nécessite ensuite la création des différentes variables binaires, discrètes et catégoriques de sorte à mener les expérimentations de règle d'association et d'algorithme de classification.

La figure ci-dessous permet de schématiser le processus de traitement de données LinkedIn.

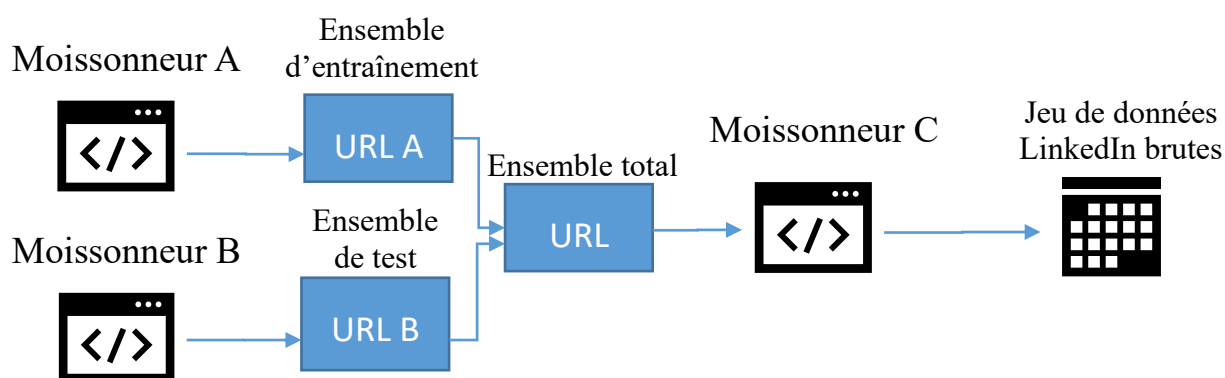


Figure 4-2. Traitement de données LinkedIn

Le Moissonneur A désigne le Moissonneur développé pour obtenir la liste d'URL des profils LinkedIn des employés de CAE Inc. qui constituent l'ensemble d'entraînement, tandis que le Moissonneur B est l'outil développé pour obtenir la liste d'URL des profils LinkedIn des employés de CAE Inc. qui constituent l'ensemble de test. Nous avons vu que les tables de sorties sont toutes deux des listes d'URL renvoyant vers les profils LinkedIn des employés, toutefois leur principe de fonctionnement n'est pas identique d'où l'importance d'avoir deux Moissonneurs distincts dont l'objectif est d'obtenir l'étape intermédiaire de la liste d'URL qui constitue notre ensemble total. Enfin, le Moissonneur C désigne l'outil qui permet d'extraire les données du CH de chaque profil LinkedIn à partir de la liste d'URL totale. Le Moissonneur C permet ainsi d'obtenir une table de données brute du CH à partir des profils LinkedIn de notre ensemble.

4.4 Construction des variables

La section sur la création de variables présente la démarche de construction des différentes variables utilisées lors des 3 modèles de l'étude exploratoire. Il s'agit de décrire la création des variables binaires, des variables discrètes et des variables catégoriques que nous utilisons en vue d'appliquer les méthodes d'apprentissage indiquées à la section 4.5. La raison qui nous a amené à créer ces variables sera justifiée lors du Chapitre 5 sur les Résultats des expérimentations et lors du Chapitre 6 sur l'Interprétation de ces résultats. Nous distinguons deux ensembles de variables, celles associées aux données brevets et celles associées aux données LinkedIn.

4.4.1 Construction des variables de données de brevets

Dans cette section, nous présentons la construction des variables de données brevets issus de la collecte de données des brevets de l'entreprise CAE Inc. sur la plateforme Patseer. D'après la section 4.1.1, sur l'extraction d'information des données brevets, nous utilisons essentiellement les domaines technologiques des brevets qui sont établis par le logiciel Patseer. Ces domaines technologiques (*Tech Domain*) et sous-technologiques (*Tech Sub Domain*) servent ensuite à créer les variables discrètes sur le nombre de brevets réalisés par l'inventeur, le nombre de brevets dans le domaine technologique favoris ainsi que le descripteur sur le ratio nombre de brevets dans le domaine technologique favoris sur le nombre de brevets total de l'inventeur.

Dans un premier temps, il s'agit de connaître les *Tech Domain* et *Tech Sub Domain* présent dans le jeu de données brevets de notre ensemble afin de pouvoir utiliser ces catégories pour la création des variables discrètes sur le nombre de brevets dans le domaine technologique « favoris » et le descripteur du ratio entre le nombre de brevets du domaine technologique favoris et le nombre de brevets déposés par l'inventeur. Les catégories *Tech Domain* et *Tech Sub Domain* sont établies par le logiciel Patseer, ces catégories sont créées à partir des codes IPC des brevets dont la classification se trouve sur le site Web de la WIPO⁹. La création des variables *Tech Domain* et *Tech Sub Domain* globale est disponible à la demande au service de logiciel Patseer. Or nous présentons dans les

⁹ La classification internationale des brevets se trouve sur le site Internet suivant : [Lien vers la classification WIPO](#)

tableaux suivants, uniquement la création des catégories pour les *Tech Domain* et *Tech Sub Domain* qui concernent les brevets réalisés par l'entreprise CAE Inc. D'une part, le tableau ci-dessous présente l'association des codes IPC des brevets pour chaque catégorie de sous-domaine technologique des brevets réalisés par CAE Inc.

Tableau 4-4. Construction des catégories de "sous-domaines technologiques" des brevets

Codage catégories ‘Tech Sub-domain’ (Patseer) avec les 4 premiers caractères des codes IPC Brevets									
OPTICS	G02B								
AUDIO-VISUAL TECHNOLOGY	H04N	H04R							
CONTROL	G09B	G05D							
MEASUREMENT	G04F	G08G	G01R	G01S	G04G	G01N	G01L	G01B	G01C
COMPUTER TECHNOLOGY	G06T	G06F							
TELECOMMUNICATIONS	H04H	H04B							
OTHER CONSUMER GOODS	G10K								
CHEMICAL ENGINEERING	B01D	B07C	B05B						
MACHINE TOOLS	B23Q	B24C	B27B						
BASIC COMMUNICATION PROCESSES	H03G								
DIGITAL COMMUNICATION	H04L	H04W							
IT METHODS FOR MANAGEMENT	G06Q	G08G							
HANDLING	B65G								
MECHANICAL ELEMENTS	F16F								
MEDICAL TECHNOLOGY	A61B	A61F							
ELECTRICAL MACHINERY - APPARATUS	F21V	F21Y	H01R						
OTHER SPECIAL MACHINES	B29C								

On retrouve les sous-domaines technologiques dans la colonne de gauche, des brevets réalisés par CAE Inc. et dans la colonne de droit les codes IPC des brevets qui sont associés à ces catégories de sous-domaines technologiques.

D'autre part, le tableau suivant décrit l'association des sous-domaines technologiques aux domaines technologiques des brevets. Ici encore nous présentons dans le tableau uniquement les catégories de domaines technologiques des brevets de l'entreprise CAE Inc. L'association des catégories complètes peut être obtenue en contactant directement le service de logiciel Patseer.

Tableau 4-5. Catégories des sous-domaines technologiques par domaines technologiques des brevets

ELECTRICAL ENGINEERING; INSTRUMENTS	ELECTRICAL ENGINEERING	INSTRUMENTS	MECHANICAL ENGINEERING; INSTRUMENTS
OPTICS ; AUDIO-VISUAL TECHNOLOGY	AUDIO-VISUAL TECHNOLOGY; COMPUTER TECHNOLOGY	CONTROL	TRANSPORT; CONTROL
AUDIO-VISUAL TECHNOLOGY ; OPTICS	AUDIO-VISUAL TECHNOLOGY; COMPUTER TECHNOLOGY; TELECOMMUNICATIONS	MEASUREMENT	MECHANICAL ELEMENTS ; CONTROL
COMPUTER TECHNOLOGY ; CONTROL	COMPUTER TECHNOLOGY	MEASUREMENT ; CONTROL	OTHER SPECIAL MACHINES ; OPTICS
COMPUTER TECHNOLOGY ; MEASUREMENT ; CONTROL	AUDIO-VISUAL TECHNOLOGY	CONTROL ; MEASUREMENT	OTHER SPECIAL MACHINES ; OPTICS
AUDIO-VISUAL TECHNOLOGY; OPTICS; COMPUTER TECHNOLOGY; CONTROL; TELECOMMUNICATIONS	COMPUTER TECHNOLOGY ; DIGITAL COMMUNICATION	CONTROL ; OPTICS	OTHER SPECIAL MACHINES ; OPTICS
CONTROL ; COMPUTER TECHNOLOGY	COMPUTER TECHNOLOGY; AUDIO-VISUAL TECHNOLOGY	CONTROL ; MEDICAL TECHNOLOGY	OTHER SPECIAL MACHINES ; OPTICS
CONTROL ; DIGITAL COMMUNICATION	COMPUTER TECHNOLOGY; IT METHODS FOR MANAGEMENT	MEASUREMENT ; OPTICS	
AUDIO-VISUAL TECHNOLOGY ; CONTROL	IT METHODS FOR MANAGEMENT	OPTICS	
COMPUTER TECHNOLOGY; IT METHODS FOR MANAGEMENT; CONTROL	ELECTRICAL MACHINERY; APPARATUS; ENERGY; DIGITAL COMMUNICATION	OPTICS ; CONTROL	
COMPUTER TECHNOLOGY ; CONTROL ; DIGITAL COMMUNICATION	COMPUTER TECHNOLOGY; BASIC COMMUNICATION PROCESSES; AUDIO-VISUAL TECHNOLOGY		
COMPUTER TECHNOLOGY; IT METHODS FOR MANAGEMENT; CONTROL; DIGITAL COMMUNICATION	ELECTRICAL MACHINERY ; APPARATUS ; ENERGY		
IT METHODS FOR MANAGEMENT; MEASUREMENT; CONTROL			
COMPUTER TECHNOLOGY; CONTROL; AUDIO-VISUAL TECHNOLOGY			
AUDIO-VISUAL TECHNOLOGY; OPTICS; COMPUTER TECHNOLOGY			
CONTROL ; MEASUREMENT ; COMPUTER TECHNOLOGY			
CONTROL ; COMPUTER TECHNOLOGY ; DIGITAL COMMUNICATION			
COMPUTER TECHNOLOGY; ELECTRICAL MACHINERY; APPARATUS; ENERGY; OPTICS; CONTROL			
ELECTRICAL MACHINERY; APPARATUS; ENERGY; OPTICS; CONTROL			
COMPUTER TECHNOLOGY ; OPTICS			
OPTICS; COMPUTER TECHNOLOGY; AUDIO-VISUAL TECHNOLOGY; TELECOMMUNICATIONS			
CONTROL; IT METHODS FOR MANAGEMENT			
CONTROL; IT METHODS FOR MANAGEMENT; DIGITAL COMMUNICATION			
DIGITAL COMMUNICATION ; COMPUTER TECHNOLOGY ; CONTROL ; AUDIO-VISUAL TECHNOLOGY			

Le tableau présente les 4 grandes catégories de domaines technologiques des brevets réalisés par CAE Inc., à savoir « ELECTRICAL ENGINEERING ; INSTRUMENTS », « ELECTRICAL ENGINEERING », « INSTRUMENTS », MECHANICAL ENGINEERING. INSTRUMENTS ». Chaque brevet possède un ou plusieurs sous-domaine(s) technologique(s) selon les codes IPC qu'il possède. C'est pourquoi chaque catégorie de domaine technologique inclue plusieurs combinaisons de sous-domaines technologiques. Ce prérequis sur les domaines technologies nous permet alors de créer les variables discrètes sur le nombre de brevet, le nombre de brevet dans le domaine favori de l'inventeur et le descripteur du potentiel d'innovation du CO de l'entreprise.

- Variable discrète du « Nombre de brevets »

La variable discrète du Nombre de brevets, comme son nom l'indique, comptabilise le nombre de brevets réalisés par l'inventeur avec l'entreprise CAE Inc. Cette variable est construite uniquement pour l'ensemble d'apprentissage, nommé ensemble d'entraînement.

- Variable discrète du « Nombre de brevets dans le domaine technologique favoris »

La variable du nombre de brevet par domaine technologique favoris est construite en dénombre le nombre de brevets réalisés par domaine technologique pour chaque inventeur. Ensuite, le domaine technologique qui comptabilise le plus de brevets, cette valeur discrète constitue ainsi la variable du nombre de brevets dans le domaine technologique dit favoris.

- Descripteur du potentiel d'innovation du CO

Le descripteur du potentiel d'innovation du CO est défini comme le ratio entre le nombre de brevets dans le domaine technologique favoris et le nombre de brevets total réalisé par l'inventeur. Ce descripteur est utilisé pour interpréter les résultats des algorithmes d'apprentissage.

4.4.2 Construction des variables de données LinkedIn

Cette section présente la création des variables binaires, discrètes et catégoriques pour les données du CH provenant de la source LinkedIn.

- **Construction des variables « Formation »**

Selon les 3 modèles présentés à la section 3.3, nous utilisons des variables que nous regroupons selon la dimension de « Formation » pour le concept du CH. Comme présenté dans la section sur

le traitement des données de LinkedIn, nous obtenons des variables brutes sur l'université, le niveau d'étude, le domaine d'étude ainsi que les associations de l'individu. Nous pouvons construire des variables à partir de ces variables brutes en respectant la méthodologie présentée ci-après afin de pouvoir appliquer les méthodes d'analyse de données.

- *Construction des variables binaires « Université »*

Compte tenu que lors de la méthode de sélection de données adoptée et décrite à la section 4.1.2, les individus qui composent notre ensemble sont localisés au Canada. La grande majorité des individus ont réalisé leurs études dans un établissement Canadien. C'est pourquoi nous décidons de créer 8 variables binaires qui prennent la valeur 0 ou 1, en fonction que l'individu ait étudié et obtenu un diplôme dans l'un des établissements suivants : **Sherbrooke, McGill, ETS, Poly, Concordia, École d'ingénieurs française, Collège canadien, Lycée Français**. Les variables se construisent à partir des 3 variables brutes « Nom d'établissement », et prennent la valeur 1 si l'individu a étudié au moins une fois dans l'établissement parmi les 3 variables et 0 sinon.

- *Construction des variables binaires « Niveau d'étude »*

Les variables binaires de Niveau d'étude sont construites à partir des grades universitaires au Canada¹⁰. Les grades conservés sont ceux les plus fréquents parmi une population d'employés de statut cadre en entreprise industrielle. Ainsi, nous construisons 4 variables binaires de Niveau d'études qui sont **PhD, Master, Baccalauréat, Technique** (diplôme d'étude collégial DEC, diplôme français correspondant à Bac +2 ex : DUT, certification). A partir du jeu de données brutes, les variables binaires sont construites en prenant la valeur 1 si l'une des 3 variables de niveau de diplôme du jeu de données brutes correspond au niveau Technique, Baccalauréat, Master (Maîtrise) ou PhD (Doctorat) et 0 sinon.

- *Construction des variables binaires « Domaine d'études »*

Les variables binaires de domaines d'études, comme leurs noms l'indiquent, représentent le ou les domaine(s) d'études de chaque employé de l'entreprise. Comme il s'agit d'une entreprise

¹⁰ La liste des grades universitaires du Canada se trouve sur le site Web suivant [ici](#).

industrielle, en l'occurrence CAE Inc. qui est une entreprise du secteur de l'aéronautique et de la santé il s'agit de représenter essentiellement les domaines d'études appropriés aux activités de l'entreprise tels que le domaine Aéronautique & Mécanique, le domaine Électronique, le domaine Informatique & Logiciel, le domaine Commercial – Management – Finance et les domaines « Autres ». Le tableau ci-dessous précise les sujets d'études du jeu de données brutes qui font prendre la valeur 1 à chaque variable binaire de domaines d'études.

Tableau 4-6. Construction des variables binaires de Domaine d'études

Variable binaire de domaine d'études	Valeurs du champ de texte sujet d'études dans le jeu de données brutes
Études Aéronautique & Mécanique	{'Génie aérospatiale';' Aerospace Engineering'; 'Aerodynamics'; 'Aeronautic' ;'Aeronautical' ;'Aeronautical Engineering' ;'Aéronautique' ;'Avionics' ; 'Astronautical Engineering' ; 'Génie mécanique' ; 'Mechanical Engineering'}
Études Électronique	{'Electrical'; 'Electrical Engineering'; 'Electronics'; 'Communications Engineering'; 'Electronic, Automation, Software'; 'Telecommunications'; 'Génie Electrique'; 'Image processing and multimedia';'imaging sciences'}
Études Informatique & Logiciel	{'Computer Engineering'; 'Computer programming';'computer science'; 'computer software'; 'computer systems'; 'Génie informatique'; 'Génie Logiciel'; 'Mobile Systems security';'multimedia concentration'; 'mathématiques et informatique' ;'software engineering'; 'System reliability and safety'; 'Human factors'}
Études Business-Management-Finance	{'Accounting' ;'Administration des affaires';'Achats, Approvisionnements' ;'ADM-gestion des opérations et stratégie' ;'Administration / Business Intelligence Strategy' ; 'Gestion des affaires' ; 'Business Administration' ; 'Management' ; 'Management de projet' ; 'Real Estate Investment & Finance' ; 'Project management' ;'Supply chain'}
Études « Autres »	Ceux n'indiquant rien sur le sujet de diplôme ou {'génie industriel', 'maths', 'physics', 'biomedical', 'medecine'}

On remarque que dans la construction des variables, il faut prendre en considération les différentes langues des profil LinkedIn notamment en Anglais et en Français dans ce cas, ainsi que les possibles fautes d'orthographes lors de la complétion du profil par l'utilisateur.

- *Construction de la variable catégorique « Domaine d'études »*

D'après la construction des variables binaires précédentes des domaines d'études, il est possible de construire une variable catégorique à 4 valeurs qui comprend toutes les combinaisons existantes dans le jeu de données des domaines d'études. Le but de cette variable catégorique est de réduire le nombre de descripteurs utilisés pour les méthodes d'apprentissage classification, développée à la section 4.5.3. En effet, utiliser 5 variables binaires avec une méthode d'apprentissage affecte le « poids » accordés aux variables lors de la recherche des classes dans l'ensemble d'individus. Le tableau ci-dessous détaille la construction de la variable catégorique.

Tableau 4-7. Construction de la variable catégorique "Domaine d'études"

Variable Domaine d'Études	
Catégorie	Valeur précédente
A	1- Génie Aéronautique et Mécanique
	7- Génie Aéronautique ET Business- Management – Finance (1+3)
B	2 – Génie Électrique
	9- Génie Électrique ET Business – Management – Finance (2+3)
C	4- Génie Informatique & Logiciel
	11- Business – Management – Finance ET Génie Informatique et Logiciel (3+4)
D	3 - Business – Management – Finance
	5- Études Autres
	6- Génie Aéronautique ET Génie Électrique (1+2)
	8- Génie Aéronautique ET Génie Informatique & Logiciel (1+4)
	9- Génie Électrique ET Business – Management – Finance (2+3)
	10- Génie Électrique ET Génie informatique & Logiciel (2+4)
	12- Génie Aéronautique ET Génie Électrique ET Business – Management – Finance (1+2+3)
	13- Génie Électrique ET Business – Management – Finance ET Génie Informatique & Logiciel (2+3+4)
	14- Génie Aéronautique ET Business – Management – Finance ET Génie Informatique & Logiciel (1+3+4)
	15- Génie Aéronautique ET Génie Électrique ET Génie Informatique & Logiciel (1+2+4)

Le choix qui a été effectué dans la construction de la variable catégorique est de regrouper les compositions de domaines d'études suivant les parcours effectués par les employés de notre ensemble. La catégorie A comprend le sujet d'études Aéronautique & Mécanique ainsi que sa composition avec le domaine d'études Business-Management-Finance (BMF). La catégorie B comprend le sujet d'études Électrique et sa composition avec le sujet d'études BMF. De même pour la catégorie C pour le sujet d'études Informatique & Logiciel et sa composition avec le sujet d'études BMF. Enfin la catégorie D comprend toutes les autres compositions de sujets d'études.

- *Construction de la variable catégorique « Diversité des domaines d'études »*

La variable catégorique sur la Diversité des domaines d'études est utilisée dans le but de mesurer les différences de parcours au sein de l'ensemble. Elle est construite selon que l'individu ait étudié dans des domaines différents au long de son parcours étudiant ou bien durant son parcours professionnel. Le tableau ci-dessous détaille la construction de la variable catégorique sur les 2 valeurs.

Tableau 4-8. Construction de la variable catégorique "Diversité des domaines d'études"

Variable Diversité des Études	
Catégorie	Valeur précédente
A	1-1 domaine de génie OU Business – Management – Finance
B	2 – Études Autres
	3 – 2 ou + diplômes en génie seulement
	4- 1 diplôme de génie ET Business – Management – Finance
	5 – 2 domaines de Génie ET Business – Management – Finance

La catégorie A inclue tous les individus qui ont étudié un seul domaine d'études en génie ou dans le domaine BMF. La catégorie B comprend les individus ayant étudiés plusieurs domaines d'études qui justifie l'attribution d'une autre catégorie de diversité des études.

- *Construction de la variable binaire « Formation continue »*

La variable binaire de « Formation continue » détermine si l'employé de l'entreprise a suivi des formations supplémentaires depuis le début de sa carrière professionnelle. Elle est calculé à partir d'une comparaison entre la date d'obtention du dernier diplôme dans les variables brutes de formation et les dates d'occupation des postes. La variable « Formation continue » prend la valeur 1 si la date d'obtention du dernier diplôme est strictement plus récente (supérieure) que les dates d'occupation des postes., et la valeur 0 sinon.

- *Construction des variables binaires « Association »*

Les variables binaires « Association » indique si l'employé a participé à une ou plusieurs associations lors de ses formations. On construit alors 3 variable binaire : Association de sport d'équipe, Association aéronautique et Association étudiante.

- **Construction des variables « Compétences »**

Cette section porte sur la construction des variables de « Compétences » en regard des groupes de compétences utiles aux professions des employés de l'entreprise CAE Inc. trouvée dans la littérature. Le groupe de variables « Compétences » inclue des variables binaires sur les groupes de compétences et une variable catégorique sur la diversité de la catégorie des compétences.

- *Construction des variables binaires « Catégories de compétences »*

La création des variables binaires sur les groupes de compétences sont construites à partir des compétences indiquées sur les profils LinkedIn des employés et que l'on retrouve à la fois dans les données brutes « Top Skills » et « Other Skills ». Le tableau ci-dessous indique les catégories de variables créées à partir des compétences individuelles.

Tableau 4-9. Construction de la variable "Catégorie de compétences"

Groupe de compétences	Compétences
Développement Informatique	{Software Development; Software as a Service (SaaS); Software Design; C++ ; C ; Matlab ; Visual Studio; .NET; Python; Programming}
Logiciels	{Software Engineering; Testing; Simulations; 3D Modeling; VBA}
Management	{Project Management; Agile Methodologies; Leadership; Management}
Relationnel	{Teamwork; Communication; Communication Skills; Business Networking; Speaking}
Bureautique	{Microsoft Office ; Microsoft Excel ; Latex ; Cisco}
Connaissances secteur aéronautique	{Aircraft; Aerospace; Aviation; Avionics; Aeronautics; Aerodynamics}

Le tableau précise les définitions des 6 groupes de compétences à partir des compétences. Les 6 variables binaires sont le Développement Informatique, Logiciel, Management, Bureautique et les Connaissances du secteur aéronautique.

- *Construction de la variable catégorique « Diversité des catégories des compétences »*

En suivant le même principe que pour les formations, nous décidons de créer une variable catégorique sur la Diversité des catégories de compétences. Le tableau ci-dessous détaille la construction de la variable catégorique.

Tableau 4-10. Construction de la variable catégorique "Diversité des catégories de compétences"

Variable Diversité des Compétences	
Catégorie	Combinaison de compétences
A	0 - Rien
B	1 – 1 catégorie parmi Secteur aéro – catégories techniques (Dev. Informatique – Logiciels d'ingénierie) – Management – Relationnel
C	2 – 2 catégories techniques
	3 – Secteur aéronautique + 1 catégorie technique OU Management + Relationnel
	4- Secteur aéronautique + 2 catégories techniques
D	5 – 1 catégorie technique + Relationnel OU 1 catégorie technique + Management
	6- Secteur aéronautique + Relationnel OU Secteur aéronautique + Management
	7- 2 catégories techniques + Relationnel OU 2 catégories techniques + Management
	8- Secteur aéronautique + 1 catégorie technique + Relationnel OU Secteur aéronautique + 1 catégorie technique + Management
	9- Secteur aéronautique + 1 catégorie technique + Relationnel OU 1 catégorie technique + Management + Relationnel
E	10 – Secteur aéronautique + 2 catégories technique + Management OU Secteur aéronautique + 2 catégories technique + Relationnel
	11- 2 catégories technique + Relationnel + Management
	12- Secteur aéronautique + 1 catégorie technique + Management + Relationnel
	13- Secteur aéronautique + 2 catégories techniques + Management + Relationnel

Le tableau présente la façon dont la variable catégorique est construite, avec une logique de diversité « croissante » de la catégorie A vers la catégorie E. En effet, la catégorie A correspond au minimum de diversité de groupe de compétences, en l'occurrence aucune tandis que la catégorie E correspond à une forte de diversité avec plusieurs combinaisons de groupes de compétences. Cette variable est utilisée pour les méthodes d'algorithmes d'apprentissage.

4.5 Modèle d'analyses

Dans cette section de méthode d'analyse des données, nous présentons les différentes méthodes et leur théories Mathématiques qui ont été appliquées à notre ensemble de données décrits à travers

les 3 modèles détaillés à la section 3.3. Les méthodes de règles d'associations, d'arbre de décision et algorithme de classifications des K-moyennes ont été choisis de par leur simplicité et leur efficacité pour extraire des connaissances grâce aux puissances de calcul informatique. Ces méthodes ont pour objectif de rechercher à partir des données disponibles un modèle explicatif ou prédictif entre, d'une part un attribut particulier à prédire, qui est ici le potentiel d'innovation du CO de l'entreprise et, d'autre part, des attributs prédictifs c'est-à-dire savoir quels attributs du CH sont déterminants dans la prédiction du potentiel d'innovation du CO. Dans le cas où un tel modèle serait produit et qu'il s'avérerait valide, il pourrait alors être utilisé à des fins de prédiction. Ceci fait l'objet d'analyse des résultats d'expérimentation dans le Chapitre 5.

En apprentissage supervisé, la recherche des attributs prédictifs s'effectue dans un premier temps pendant **une phase « inductive »** consistant à développer les règles d'identification à partir d'exemples particuliers et, dans un second temps la recherche de l'attribut particulier à prédire est faite durant **une phase « prédictive »** visant à utiliser ces règles pour identifier de nouvelles instances. C'est le principe de fonctionnement des méthodes de règles d'association et des arbres de décisions qui sont présentés dans cette section. Cependant, les méthodes de prédiction ne procèdent pas toutes ainsi. Les méthodes de prédiction à base d'instances comme l'algorithme des « K-moyennes » n'établissent pas une liaison fonctionnelle entre l'attribut à prédire et les valeurs des attributs prédictifs.

4.5.1 Règles d'associations

A l'origine, la méthode de règles d'association est connue sous le nom d'analyse du panier de la ménagère. Cette méthode de règle d'association était utilisée dans le domaine du marketing pour identifier les habitudes d'achats des clients, ou comportements clients, entre des ensembles de produits, des articles et des catégories de produits (Berry & Linoff, 2004). Cette méthode est alors devenue puissante pour comprendre les règles d'association des produits vendus et ainsi établir une stratégie avec la prédiction des ventes de produits en fonction des habitudes des clients.

Il existe de nombreux cas d'études d'application présents dans la littérature avec par exemple : l'utilisation des règles d'association pour connaître le comportement d'achat des consommateurs d'Amazon, ou encore l'utilisation des règles d'associations pour découvrir les allergies que les gens avaient (Kanagawa et al., 2009).

La méthode des règles d'association se présente alors comme un outil ayant la capacité de s'appliquer à différents domaines d'application, c'est pourquoi nous décidons d'utiliser cette méthode dans l'objectif d'identifier le potentiel d'innovation du CO d'une entreprise industriel à partir du CH de ses employés.

Parmi les différents algorithmes de règles d'association (A Priori, A Priori Partition, FP-Tree, FP-Growth, RARM, ARM, etc.) qui sont fréquemment appliqués lors des analyses statistiques, **nous avons opté pour l'algorithme « apriori » (Agrawal & Srikant, 2000) pour sa simplicité de mise en œuvre lors d'extraction de connaissances à partir d'un ensemble de données.** En effet, il s'agit simplement de dénombrer les combinaisons et de conserver celles qui sont les plus fréquentes, c'est-à-dire qui dépassent un certain seuil fixé dans l'ensemble de données. Il s'agit de la première méthode explicative et de prédiction appliquée sur notre jeu de données correspond au

Modèle de la

Figure 3-1. Modèle 1 de variables binaires.

La théorie suivante, présentée par la méthode documentée des Techniques de l'Ingénieur, traité Informatique dans Extraction de la connaissance à partir de données (ECD) nous permet de comprendre les principes mathématiques sous-jacents des règles d'association. De plus, nous adaptons la théorie présentée au cas d'application de notre étude exploratoire.

Les p variables statistiques $X = (X_1, X_2, \dots, X_p)$ sont toutes booléennes et sont appelées « items ».

Les items représentent alors les variables binaires qui sont précisées dans la

Figure 3-1. Modèle 1 de variables binaires. La population Ω est composée de l'ensemble des transactions (où la transaction désigne un ensemble de variables possibles). Comme toutes les variables sont booléennes, une transaction ω est caractérisée par un sous-ensemble de variables $X' \subseteq X$ (ou ensemble d'items) toutes à 1.

Soit X_a un ensemble d'items, $X_a \subseteq X$; une transaction ω contient l'ensemble d'items X_a si et seulement si tous les items qui forment X_a sont à 1 dans la transaction ω . Par exemple, si X_a désigne quatre produits du magasin que l'on note A, B, C et D, on dira que X_a est contenu dans la transaction ω si les quatre produits figurent dans la transaction, on écrira alors $X_a \subseteq \omega$.

Une règle d'association est une implication de la forme $X_a \Rightarrow X_b$ où $X_a \subset X$ et $X_b \subset X$ et $X_a \cap X_b = \emptyset$. La règle $X_a \Rightarrow X_b$ se produit dans l'ensemble des transactions Ω avec un support s qui est le

pourcentage de transactions de Ω qui contient $X_a \cup X_b$ et une confiance c qui est le pourcentage de transactions Ω contenant X_a qui contiennent également X_b . Ainsi :

$$\begin{aligned} support(X_a \Rightarrow X_b) &= P(X_a \cup X_b) = s \text{ et} \\ confidence(X_a \Rightarrow X_b) &= P(X_b/X_a) = c \end{aligned}$$

Les règles de la forme $X_a \Rightarrow X_b$ sont dites pertinentes si $support(X_a \Rightarrow X_b) > s_0$ et $confidence(X_a \Rightarrow X_b) > c_0$ où s_0 et c_0 sont des seuils minimums.

L'algorithme de base Apriori préconise la recherche des ensembles fréquents de cardinal $k + 1$ à partir des ensembles fréquents de cardinal k . Ainsi, pour trouver les ensembles fréquents ayant deux items, on utilisera exclusivement les ensembles fréquents ayant un item. Il est clair qu'en toute généralité, le nombre d'ensembles fréquents diminue avec le nombre d'items : il y a moins d'ensembles fréquents à deux items qu'à un item. Cette propriété permet ainsi de restreindre la taille de l'espace à explorer pour trouver tous les ensembles fréquents nécessaires à la deuxième étape de l'algorithme qui comporte deux points :

- Pour chaque ensemble fréquent X_a , on génère tous les sous-ensembles non vides ;
- Pour chaque sous-ensemble non vide $X_b \subset X_a$, on génère la règle :

$$(X_b \Rightarrow (X_a - X_b)) \text{ si } \frac{support(X_a)}{support(X_b)} > c_0$$

De plus, nous utiliserons un paramètre nommé « *lift* » qui est défini par Brin et al. (1997), par l'expression suivante :

$$lift(X_a \Rightarrow X_b) = \frac{confidence(X_a \Rightarrow X_b)}{support(X_b)}$$

Ainsi qu'un paramètre nommé « conviction » qui est également introduit par Brin et al. (1997) avec l'expression suivante :

$$conviction(X_a \Rightarrow X_b) = \frac{1 - support(X_b)}{1 - confidence(X_a \Rightarrow X_b)}$$

En pratique, nous simulons cet algorithme des règles d'association Apriori avec la bibliothèque sous Python appelée « mlxtend »¹¹. Les résultats de cette expérimentation seront présentés au Chapitre 5.

4.5.2 Arbres de décision

Les méthodes d'arbre de décision sont des méthodes polythétiques. Les méthodes polythétiques recherchent des partitions, qui sont dans notre cas des groupes d'employés de l'entreprise industrielle étudiée, dans lesquelles **les éléments d'une même classe ont, entre eux, une certaine ressemblance, et des éléments appartenant à des classes différentes d'une même partition qui doivent être les plus dissemblables possibles au sens d'un certain critère préétabli**. La ressemblance doit prendre en compte la totalité des « attributs descriptifs ». On distingue les méthodes selon qu'elles conduisent à une hiérarchie de partitions emboîtées ou non, ou à une partition à un nombre de groupes prédéterminé ou non. Parmi les techniques fréquemment employées, on trouve les méthodes de classification hiérarchique, les nuées dynamiques proposée par Diday (1971), la classification non hiérarchique descendante proposée par Fages et al. (1987). On peut également incorporer les algorithmes développés dans les domaines de l'intelligence artificielle comme l'algorithme Étoile proposé par Michalski (1983) qui s'inspire de l'algorithme des nuées dynamiques ou les algorithmes COBWEB par Arifovic (1994) et AUTOCLASS par Cheeseman et al. (1988).

Nous nous intéressons donc aux algorithmes d'arbres de décision constituent une méthode explicative et prédictive d'un ensemble de données dont le but est d'estimer les valeurs « d'attributs prédictifs » (des classes de potentiels d'innovation du CO) à partir « d'attributs explicatifs » (les variables de données de notre modèle). Les arbres de décisions sont un cas précis des graphes d'induction. Les graphes d'induction sont faciles à mettre en œuvre, les résultats qu'ils fournissent sont aisés à interpréter et les modèles qui en sont déduits sont performants. Les graphes d'induction sont capables d'appréhender des bases de données de grande taille et sont applicables sans restriction à des données de n'importe quel type (qualitatives, quantitatives ou un mélange des

¹¹ La description de la bibliothèque est présentée sur le site Web suivant [ici](#) et le code Python est disponible sur la page GitHub suivante [ici](#).

deux). La prolifération des logiciels fondés sur ce paradigme est d'ailleurs un signe patent de ce succès.

Plus particulièrement nous nous intéressons aux arbres de décision selon l'algorithme de Classification Hiérarchie Ascendante (CAH). À partir de la partition grossière formée de tous les individus de l'ensemble d'apprentissage, on recherche, parmi les p variables (X_1, X_2, \dots, X_p), celle qui permet d'engendrer la meilleure partition au sens d'un critère donné. Celui-ci devra être d'autant meilleur que les classes de la partition sont homogènes. Nous obtenons un arbre à deux niveaux dont la racine représente la partition grossière Ω et dont les feuilles représentent les modalités de la variable.

Ainsi, l'objectif de l'algorithme qui est de former des classes le plus homogène possible au sein de la population des employés de l'entreprise, utilise une mesure d'entropie comme critères d'homogénéité. La mesure de l'entropie de Shannon est celle la plus adoptée. En particulier, pour un nombre de m classes déterminées et en nommant les classes c_j qui représentent le nombre d'employés dans la classe j , l'entropie de Shannon s'exprime par la formule suivante :

$$H(\Omega) = - \sum_{j=1}^m p\left(\frac{c_j}{\Omega}\right) \cdot \log_2\left(p\left(\frac{c_j}{\Omega}\right)\right)$$

Ensuite, l'importance des variables (X_1, X_2, \dots, X_p) dans le jeu de données est définie par leur gain informationnel G_1, G_2, \dots, G_p . C'est un moyen de mesurer l'importance de chaque variable. On cherchera la variable qui réduit l'entropie ou bien celle qui maximise le gain informationnel.

L'arbre de décision ci-dessous illustre ces critères de classification à partir de la mesure de l'entropie choisie :

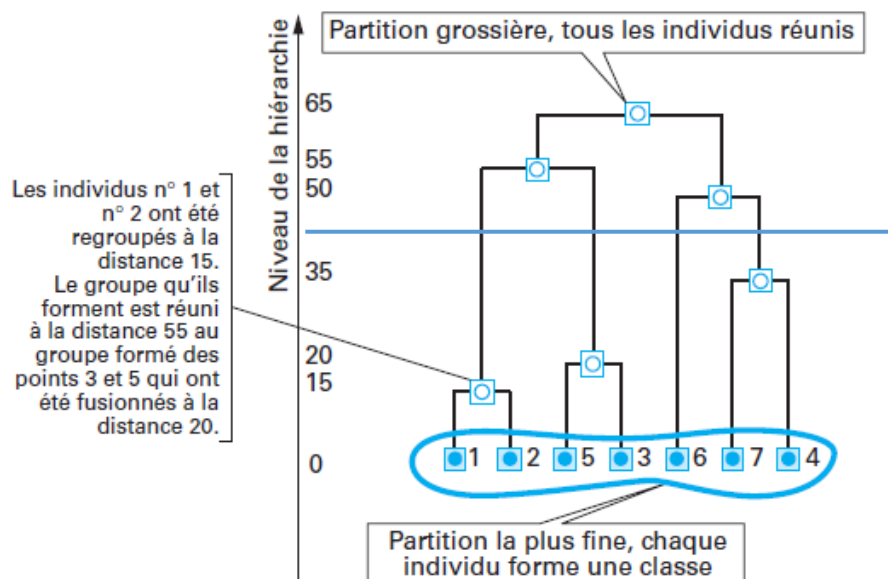


Figure 4-3. Exemple d'arbre de décision de Classification Ascendante Hiérarchique

Sur cette figure illustrative on remarque que le niveau de la hiérarchie est défini par la mesure de l'entropie, qui dans notre cas est la formule de l'entropie de Shannon. Le regroupement d'individus de la population globale (située au bas de l'axe) s'effectue à certains « niveaux de hiérarchie » qui correspondent une certaine valeur d'entropie. Plus la valeur de cette mesure est faible plus les individus du même ensemble forment un groupe homogène. Il convient alors de fixer un seuil sur ces valeurs de niveaux de hiérarchie afin de former le nombre m de classes sur la population Ω . Ici sur la Figure 4-3, le seuil fixé à la valeur 40 permet de former 4 classes composées des individus $\{1;2\}$, $\{5;3\}$, $\{6\}$ et $\{7;4\}$. Le nombre de classes est donc défini a posteriori des calculs d'entropie entre les sous-ensembles de la population.

En pratique, lors des expérimentations avec le langage Python, présentées dans le Chapitre 5, nous utiliserons la bibliothèque « `Scipy.cluster.hierarchy` »¹² pour appliquer la méthode de Classification Ascendante Hiérarchique à notre jeu de données utilisée pour les 3 modèles définis à la section 3.3.

¹² La description de la bibliothèque Python « `scipy.cluster.hierarchy` » est accessible sur le site Internet, [ici](#).

4.5.3 Algorithme des K-Moyennes

L'algorithme des K-Moyennes est un algorithme de partition de reconnaissances de forme qui permet de résoudre le même problème de prédiction précédent sans connaître au préalable la classe d'appartenance des ensembles (également appelées instances) de notre ensemble de données. L'algorithme des K-Moyennes procède donc de façon exploratoire sur l'ensemble constitué de n observations, dont chacune comporte p variables (également nommé descripteurs), dont l'objectif est de regrouper ces n observations en K classes. Le regroupement de ces observations en classes se fait grâce à des mesures de similarité et dissimilarité entre les observations par l'usage d'une fonction distance d . La fonction distance que nous utilisons lors de l'étude exploratoire est la distance euclidienne. L'objectif de l'algorithme des K-Moyennes est alors de trouver des classes homogènes et différenciées telles que les observations soient les plus similaires possibles au sein d'une classe (compacité intra-classe) et les classes soient aussi dissemblables les unes des autres que possible (séparabilité inter-classe).

Le principe de l'algorithme des K-Moyennes est le suivant : dans un premier temps K instances sont sélectionnées aléatoirement parmi l'ensemble de données entier. Ces points représentent les K centroïdes initiaux (ou centres de gravités). En utilisant la fonction de distance d toutes les autres instances sont associées à l'un des K centroïdes pour lesquelles la distance entre l'instance et l'un des K centroïdes est minimale. Ensuite les coordonnées des centroïdes sont recalculées à chaque étape de la façon suivante : les coordonnées des nouveaux K centroïdes sont la moyenne des coordonnées de l'ensemble des instances appartenant à cette même classe à l'étape actuelle. Le processus est itératif et s'arrête lors que la fonction coût atteint un optimum, sans garantie que cette valeur soit le minimum global. Cette valeur de l'optimum atteint par la fonction coût dépend de la sélection des centroïdes initiaux dans la phase d'initialisation d'où l'importance de sélectionner le meilleur set de centroïdes initiaux (Ostrovsky & Rabani, 2006).

La fonction coût utilisée dans l'algorithme de K-Moyennes se nomme inertie totale, elle est définie par l'expression suivante :

$$I_g = \frac{1}{n} \sum_{i=1}^n d^2(x_i, g)$$

Où l'instance g est définie comme le centre de gravité de l'ensemble des données.

De plus la fonction inertie totale respecte le Théorème de Huyghens qui est le suivant :

$$I_g = \frac{1}{n} \sum_{k=1}^K n_k d^2(g_k, g) + \frac{1}{n} \sum_{k=1}^K \sum_{i \in C_k} d^2(x_i, g_k)$$

Où les termes x_i représentent les instances de l'ensemble de données.

L'expression s'interprète comme l'inertie inter-classe pour le premier terme et l'inertie intra-classe dans le second terme. L'objectif de l'algorithme des K-Moyennes est d'obtenir une faible inertie intra-classe et une forte inertie inter-classe. Une solution triviale à l'optimisation de cette fonction coût revient à choisir autant de classes K que le nombre de données n . Cependant, l'objectif de l'application de l'algorithme d'apprentissage dans notre étude exploratoire est de découvrir des similitudes entre les individus de l'ensemble de données tant que possible en vue de définir des catégories de potentiel d'innovation du CO de l'entreprise, et non pas de rester à l'unicité de l'individu.

Les résultats des expérimentations de l'algorithme K-Moyennes appliqués au Modèle 3 sont présentés dans le Chapitre 5

CHAPITRE 5 RÉSULTATS

Le chapitre 5 sur les résultats proposent de suivre les résultats d'expérimentation obtenus en suivant la méthodologie décrite au chapitre 4. Notamment, les résultats permettent d'obtenir des observations sur notre cas d'étude avec la population des employés de l'entreprise CAE Inc. afin d'évaluer les modèles choisis et les méthodes d'apprentissage envisagées pour notre jeu de données. Le chapitre comporte alors des sections de statistiques descriptives et d'analyses statistiques. Plus particulièrement les résultats seront présentés en précisant le modèle utilisé, le jeu de données brutes étant le même pour toute l'étude exploratoire. Les variables utilisées sont les mêmes que décrites durant la méthodologie à la section 4.4.

5.1 Statistiques descriptives

Dans un premier temps des expérimentations, nous avons réalisé des statistiques descriptives sur les différentes variables de nos 3 modèles. Ces statistiques descriptives permettent d'une part d'obtenir une vue d'ensemble sur les valeurs de notre jeu de données afin de s'assurer que les données d'entrées ne sont pas biaisées, d'autre part de vérifier que les variables catégoriques présentent des catégories plutôt homogènes en termes de support.

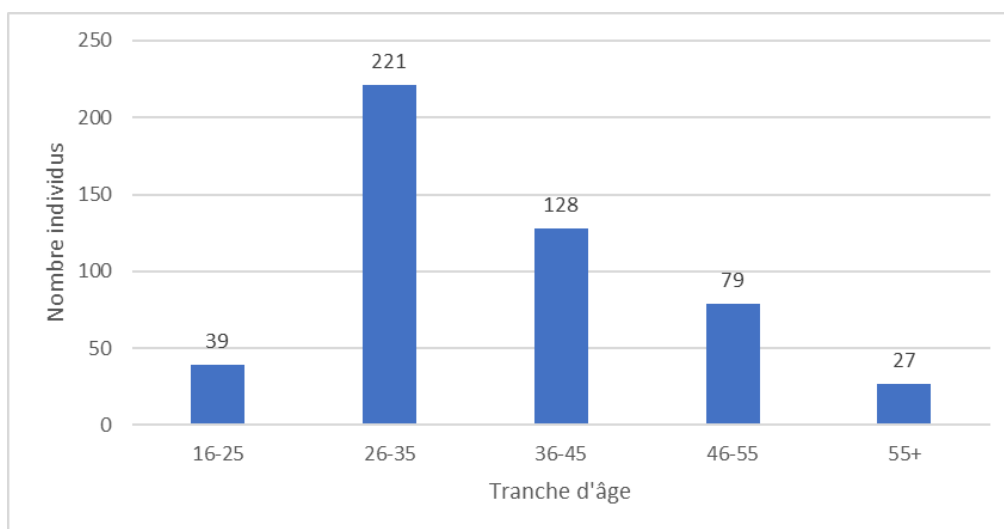


Figure 5-1. Distribution des tranches d'âge parmi la population

La population de notre ensemble total est composée principalement d'employés âgés de 26 à 35 ans avec une distribution qui semble suivre une loi normale.

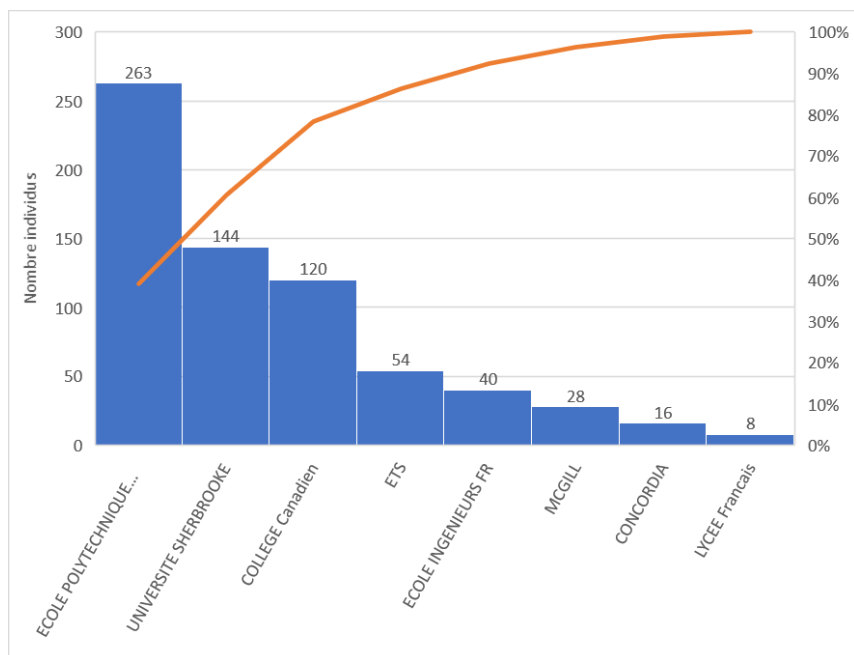


Figure 5-2. Distribution des établissements académiques fréquentés par la population

Comme nous l'avons vu dans la partie de Méthodologie sur la méthode de sélection des données, nous avons sélectionné un ensemble de test dont les employés de l'entreprise CAE Inc. étaient diplômés au moins de l'un des établissements parmi l'École Polytechnique de Montréal ou de l'Université de Sherbrooke. Il est donc normal de retrouver une distribution qui comprend environ 70% de la population parmi ces deux établissements. Par ailleurs, nous nous apercevons que des employés ont été diplômés de l'ETS, d'écoles d'ingénieurs françaises, de McGill et de Concordia ce qui implique que la variable de diversité des domaines d'études aura son importance.

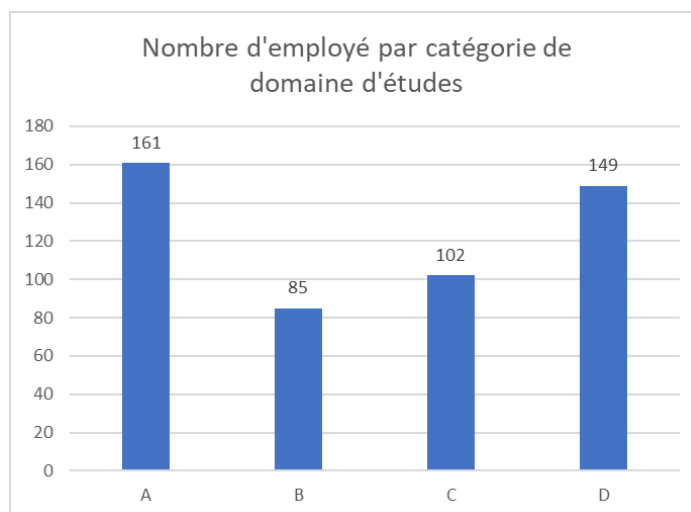


Figure 5-3. Distribution des catégories de domaines d'études

Dans la suite logique de la Figure 5-2. Distribution des établissements académiques fréquentés par la population, nous observons une distribution plutôt homogène des catégories de domaines d'études avec pour rappel « A » des études autour de l'Aéronautique & Mécanique et/ou BMF, « B » des études autour de l'Électronique et/ou BMF, « C » des études autour de l'Informatique & Logiciels et/ou BMF, et « D » des combinaisons de ces domaines d'études.

Le diagramme ci-dessous nous présente la distribution des diversités des domaines d'étude par catégorie :

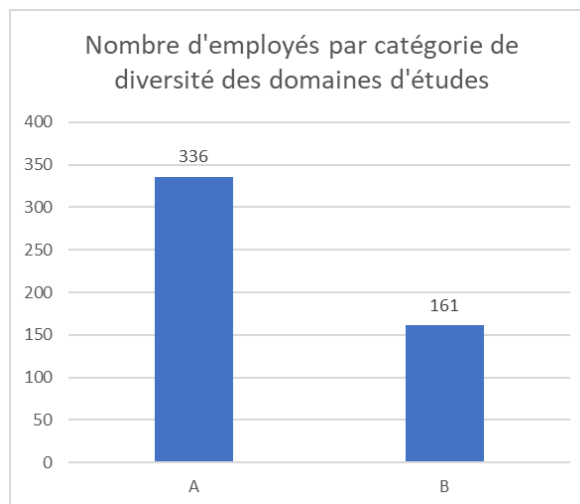


Figure 5-4. Distribution des catégories de diversité des domaines d'études parmi la population

La catégorie A correspond à la population ayant étudié dans un seul domaine d'études tandis que la catégorie B correspond à la population ayant étudié dans au moins deux domaines d'études.

Le diagramme ci-dessous nous présente la distribution des niveaux d'études parmi la population :

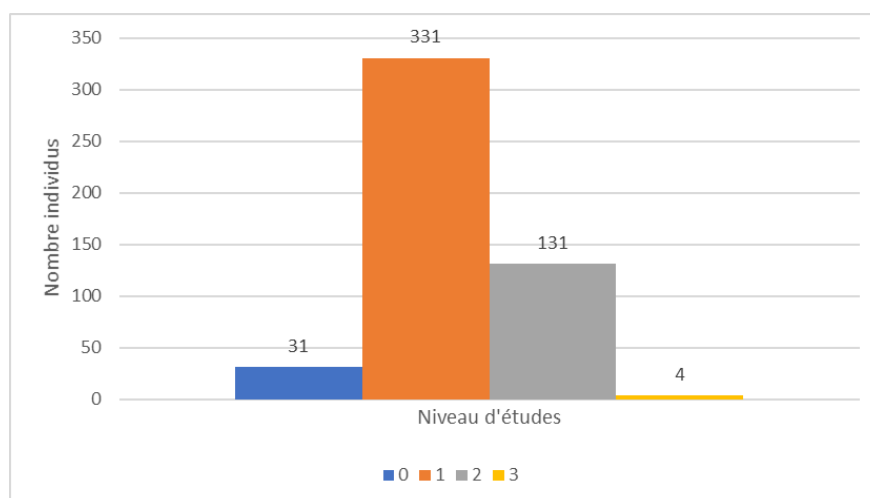


Figure 5-5. Distribution des niveaux d'études parmi la population

Ce graphique nous montre que le niveau d'études des employés de CAE Inc. de notre ensemble est principalement le niveau Bachelor (Baccalauréat) avec la valeur 1. La valeur 0 correspond à un niveau Technique uniquement, la valeur 2 correspond à un niveau Master (Maîtrise) et le niveau 3 correspond à un niveau PhD (Doctorat).

La figure ci-dessous présente le nombre d'employés qui possèdent le groupe de compétences indiquées.

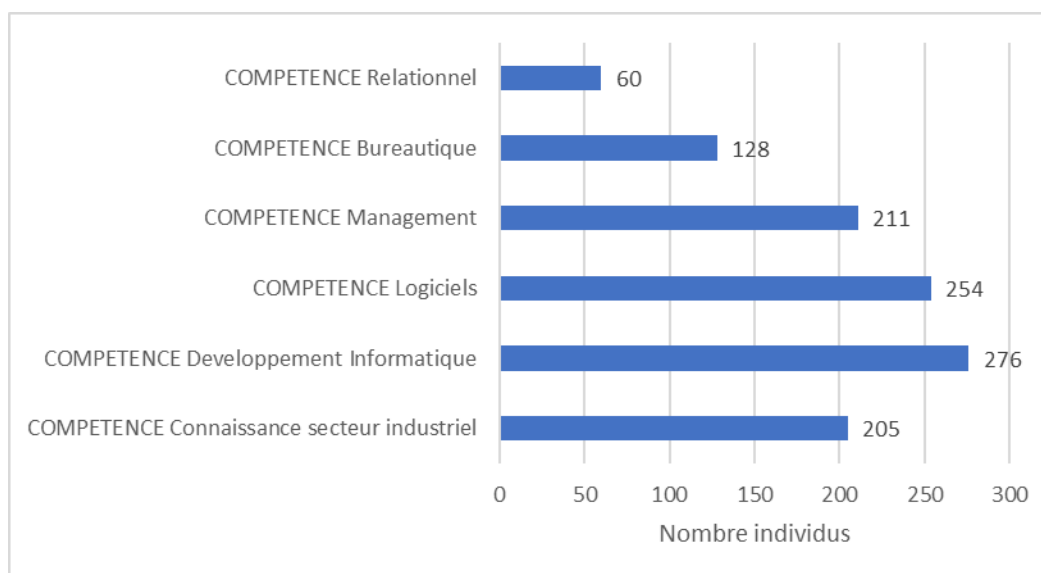


Figure 5-6. Distribution des groupes de compétences à travers la population

On relève un bon nombre d'employés qui possèdent des compétences en Développement Informatique et en Logiciels d'ingénierie à travers notre ensemble. De plus, on note que les compétences d'ordre Relationnel sont moins présentes dans le jeu de données. Ces statistiques descriptives sont peut-être représentatives des employés de l'entreprise CAE Inc. qui emploient principalement des profils techniques de l'ingénierie aéronautique et de la santé.

Le graphe suivant décrit les statistiques descriptives des catégories de diversité des groupes de compétences :

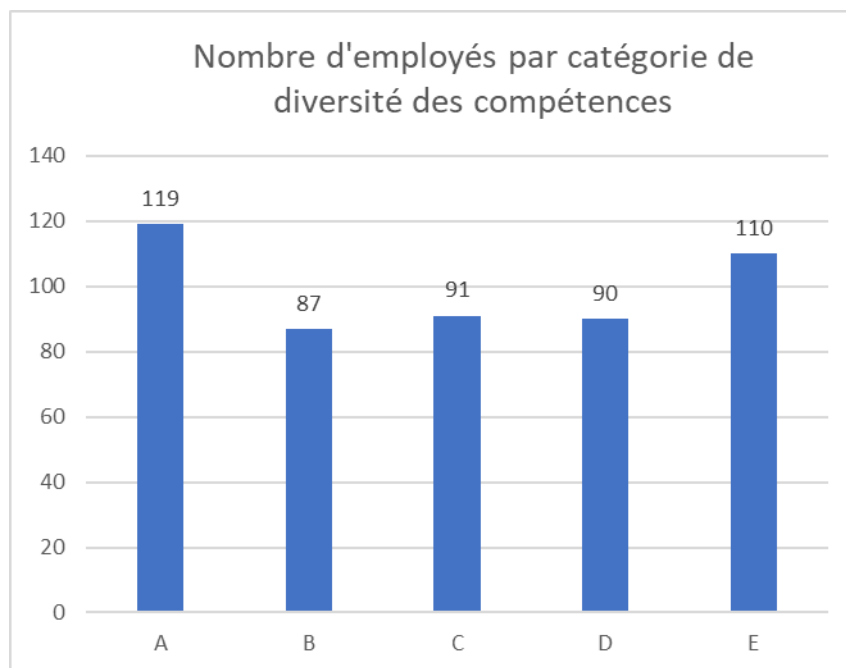


Figure 5-7. Distribution des catégories de diversité de groupes de compétence parmi la population

Encore une fois la distribution des supports (nombre d'employés appartenant à la catégorie) de chaque catégorie de diversité des groupes de compétences semble homogène. Cette distribution est importante pour valider le fait que la variable catégorique est bien construite si l'on sait que la population présente des profils suffisamment diversifiés.

Finalement nous présentons également les statistiques descriptives à propos du nombre de brevets réalisé par la population :

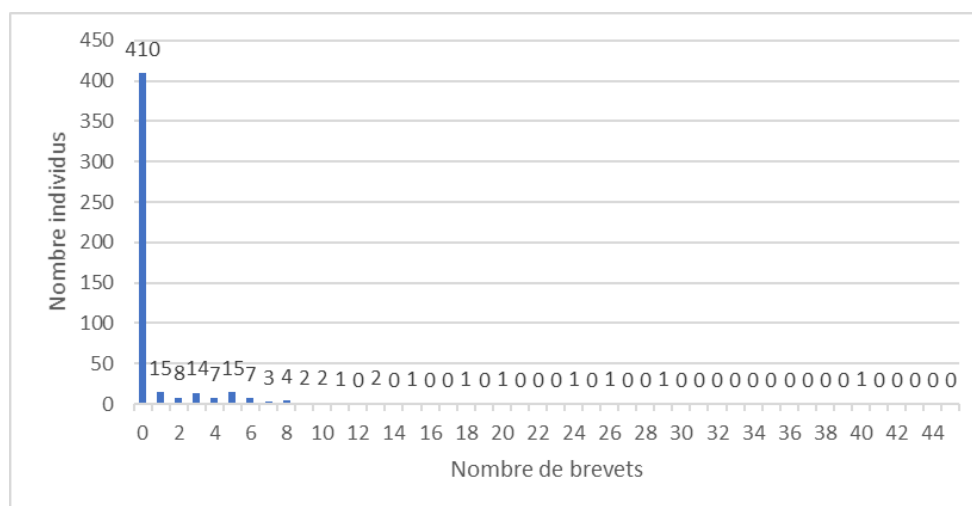


Figure 5-8. Distribution du nombre de brevets réalisés par la population

On remarque que la population n'ayant pas réalisé de brevets, qui constitue l'ensemble de test, représente 82 % de l'ensemble total. Ce qui n'est pas négligeable dans l'interprétation des résultats qui vont suivre à propos des algorithmes d'apprentissage. En effet, l'ensemble d'entraînement possède un support relativement faible comparé à l'ensemble de test pour lequel le potentiel d'innovation sera prédit.

5.2 Analyses statistiques

Cette section sur les analyses statistiques est réservée à la présentation des résultats des méthodes de règles d'association sur le modèle 1 avec l'ensemble d'entraînement, des arbres de décision sur les 3 modèles et les algorithmes de K-Moyennes sur les modèles 2 et 3. Pour rappel, les modèles utilisés sont présentés à la section 3.3.

5.2.1 Règles d'association

La méthode des règles d'association est appliquée sur le modèle 1 qui comporte uniquement des variables binaires. Dans l'expérimentation des règles d'association sur notre modèle, il est possible de fixer un seuil au nombre de règles sur certains paramètres tels que la valeur du support, la valeur de la conviction, etc. De plus, nous pouvons choisir l'ensemble d'items en sortie des règles d'association. Comme nous souhaitons obtenir les règles d'association qui mène à déterminer le potentiel d'innovation du CO avec les publications de brevets, nous choisissons l'ensemble {Brevet} en sortie des règles d'association. Nous appliquons alors la méthode des règles d'association sur le modèle 1 présenté précédemment à la Figure 3-1.

Le tableau ci-dessous présente les résultats des règles d'associations obtenues avec la bibliothèque « mlxtend.frequent_patterns » en Python.

Tableau 5-1. Résultats des règles d'association à partir du Modèle 1 de variables binaires

antecedents	consequents	antecedent support	consequent support	support	confiance	lift	leverage	conviction
{'ETUDES Informatique - logiciel', 'COMPETENCE DEV INFORMATIQUE'}	{'BREVET'}	0,1884	0,1743	0,0541	0,2872	1,6475	0,0213	1,1584
{'ETUDES Autre'}	{'BREVET'}	0,2325	0,1743	0,0661	0,2845	1,6317	0,0256	1,1539
{'ETUDES Informatique - logiciel', 'BACHELORS'}	{'BREVET'}	0,2044	0,1743	0,0501	0,2451	1,4058	0,0145	1,0937
{'ETUDES Informatique - logiciel'}	{'BREVET'}	0,2485	0,1743	0,0601	0,2419	1,3877	0,0168	1,0892
{'COMPETENCE MANAGEMENT', 'COMPETENCE LOGICIELS INGENIERIE', 'BACHELORS'}	{'BREVET'}	0,2625	0,1743	0,0621	0,2366	1,3573	0,0164	1,0816
{'COMPETENCE MANAGEMENT', 'COMPETENCE LOGICIELS INGENIERIE', 'COMPETENCE DEV INFORMATIQUE', 'BACHELORS'}	{'BREVET'}	0,2164	0,1743	0,0501	0,2315	1,3277	0,0124	1,0743
{'COMPETENCE MANAGEMENT', 'COMPETENCE DEV INFORMATIQUE', 'BACHELORS'}	{'BREVET'}	0,2505	0,1743	0,0561	0,2240	1,2848	0,0124	1,0640
{'COMPETENCE LOGICIELS INGENIERIE', 'COMPETENCE SECTEUR INDUSTRIEL', 'BACHELORS'}	{'BREVET'}	0,2425	0,1743	0,0541	0,2231	1,2799	0,0118	1,0628
{'COMPETENCE LOGICIELS INGENIERIE', 'COMPETENCE DEV INFORMATIQUE', 'BACHELORS'}	{'BREVET'}	0,3166	0,1743	0,0701	0,2215	1,2706	0,0149	1,0606
{'COMPETENCE MANAGEMENT', 'BACHELORS'}	{'BREVET'}	0,3367	0,1743	0,0741	0,2202	1,2632	0,0154	1,0589

Tableau 5-1. Résultats des règles d'association à partir du Modèle 1 de variables binaires (suite et fin)

antecedents	consequents	antecedent support	consequent support	support	confiance	lift	leverage	conviction
{'COMPETENCE LOGICIELS INGENIERIE', 'BACHELORS'}	{'BREVET'}	0,3968	0,1743	0,0842	0,2121	1,2166	0,0150	1,0479
{'COMPETENCE DEV INFORMATIQUE', 'BACHELORS'}	{'BREVET'}	0,4188	0,1743	0,0882	0,2105	1,2075	0,0152	1,0458
{'COMPETENCE MANAGEMENT', 'COMPETENCE LOGICIELS INGENIERIE', 'COMPETENCE DEV INFORMATIQUE'}	{'BREVET'}	0,2605	0,1743	0,0541	0,2077	1,1912	0,0087	1,0421
{'COMPETENCE MANAGEMENT', 'COMPETENCE DEV INFORMATIQUE'}	{'BREVET'}	0,3106	0,1743	0,0641	0,2065	1,1841	0,0100	1,0405
{'COMPETENCE MANAGEMENT', 'COMPETENCE LOGICIELS INGENIERIE'}	{'BREVET'}	0,3206	0,1743	0,0661	0,2063	1,1830	0,0102	1,0402
{'COMPETENCE MANAGEMENT'}	{'BREVET'}	0,4228	0,1743	0,0862	0,2038	1,1689	0,0124	1,0370
{'COMPETENCE LOGICIELS INGENIERIE', 'COMPETENCE SECTEUR INDUSTRIEL', 'COMPETENCE DEV INFORMATIQUE'}	{'BREVET'}	0,2585	0,1743	0,0521	0,2016	1,1560	0,0070	1,0341
{'COMPETENCE DEV INFORMATIQUE'}	{'BREVET'}	0,5551	0,1743	0,1102	0,1986	1,1388	0,0134	1,0302
{'COMPETENCE SECTEUR INDUSTRIEL', 'BACHELORS'}	{'BREVET'}	0,3046	0,1743	0,0601	0,1974	1,1320	0,0070	1,0287
{'COMPETENCE LOGICIELS INGENIERIE', 'COMPETENCE DEV INFORMATIQUE'}	{'BREVET'}	0,4088	0,1743	0,0802	0,1961	1,1246	0,0089	1,0270
{'COMPETENCE LOGICIELS INGENIERIE', 'COMPETENCE SECTEUR INDUSTRIEL'}	{'BREVET'}	0,3246	0,1743	0,0621	0,1914	1,0976	0,0055	1,0210
{'COMPETENCE SECTEUR INDUSTRIEL', 'COMPETENCE DEV INFORMATIQUE'}	{'BREVET'}	0,2846	0,1743	0,0541	0,1901	1,0906	0,0045	1,0195
{'COMPETENCE LOGICIELS INGENIERIE'}	{'BREVET'}	0,5110	0,1743	0,0962	0,1882	1,0796	0,0071	1,0171
{'BACHELORS'}	{'BREVET'}	0,7595	0,1743	0,1343	0,1768	1,0140	0,0018	1,0030
{'COMPETENCE SECTEUR INDUSTRIEL'}	{'BREVET'}	0,4108	0,1743	0,0681	0,1659	0,9513	-0,0035	0,9898

La valeur de conviction inférieure ou égale à 1 indique que l'événement contraire produit une meilleure règle. Il s'agit donc de s'assurer que les valeurs de conviction des règles analysées sont supérieures à 1.

De plus, si le paramètre du *lift* a une valeur supérieure à 1 cela indique que la probabilité de la règle *antécédent* \Rightarrow *conséquent* a plus de chance de se produire si l'antécédent et le conséquent se produisent simultanément. Si l'événement *antécédent* et *conséquent* sont indépendants alors le *lift* prendra une valeur de 1.

Ainsi en observant le tableau des résultats des règles d'association, on s'aperçoit que les employés qui ont effectué des études en Informatique & Logiciel ont tendance à publier des brevets, de part les valeurs élevées du *lift* sur les 3 premières règles (1,6475;1,6317;1,4058) et des valeurs de conviction correctes (1,1584 ; 1,1539 ;1,0937).

5.2.2 Arbre de décision du Modèle 2

Le deuxième type de méthode d'extraction de données utilisées est celle des arbres de décision. Nous avons appliqué cette méthode selon l'algorithme Apriori comme mentionné à la section 4.5.2.

L'arbre de décision ci-dessous représente les classes obtenues en fixant le seuil à la valeur de distance $d=25$ sur l'ensemble d'entraînement composé uniquement des inventeurs de brevets.

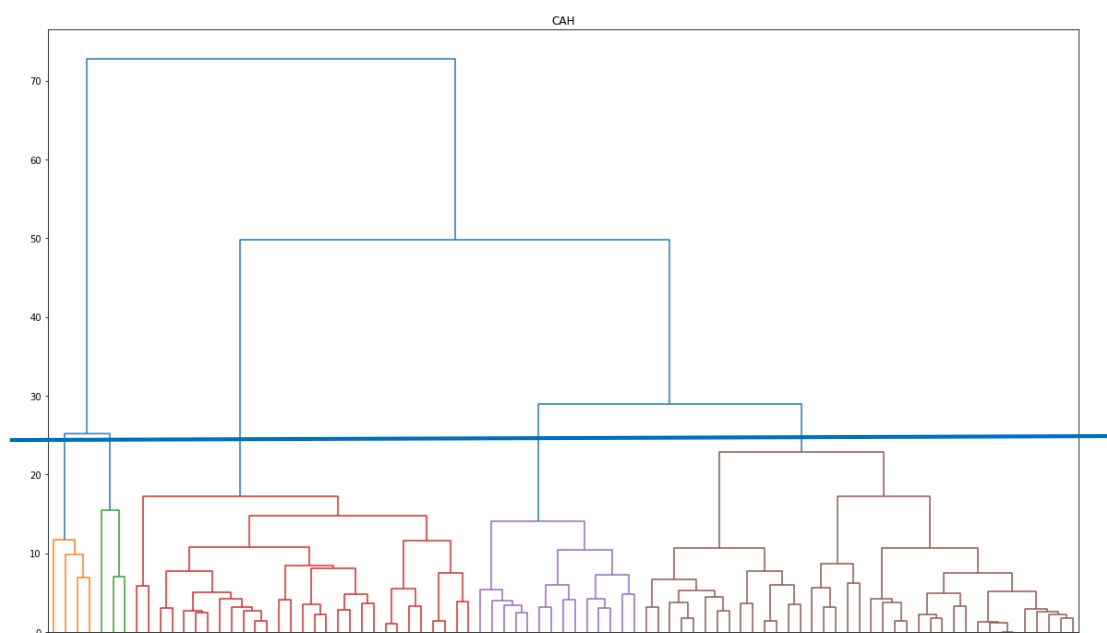


Figure 5-9. Dendrogramme (seuil = 5.0 classes)

Le nombre de classe est fixé a posteriori de la Classification Ascendante Hiérarchique et fixée à 5 par l'intermédiaire du seuil égal à 25.

Dans l'ordre de visualisation des classes obtenues sur l'ensemble d'entraînement, on retrouve les similitudes suivantes au sein des mêmes classes.

- Classe 1 : ETS
- Classe 2 : Pas de diplôme
- Classe 3 : École Polytechnique de Montréal + Baccalauréat OU Technique
- Classe 4 : École Polytechnique de Montréal + Master
- Classe 5 : Université de Sherbrooke

Pour les mêmes raisons que les résultats des règles d'association, les caractéristiques qui indiquent la classe d'appartenance en vue d'établir une règle fonctionnelle pour l'attribution d'un potentiel d'innovation du CO, ne sont pas suffisantes. Il semble naturel qu'une règle du type « Une personne qui a étudié à l'ETS, ait (ou n'ait pas) un potentiel d'innovation élevé » ne soit pas vraie de façon objective.

5.2.3 Algorithme des K-Moyenne du Modèle 3

Nous décidons d'appliquer l'algorithme des K-Moyennes sur le 3ème modèle détaillé à la Figure 3-3 qui comporte des variables binaires, discrètes et catégoriques comme elle est rappelée ci-dessous.

Le modèle comporte 3 variables catégoriques qu'il est nécessaire de transformer avant d'appliquer l'algorithme des K-Moyennes sur le jeu de données. En effet, l'algorithme des K-Moyenne tel qu'il est programmé en Python sous la bibliothèque « *sklearn.cluster* » ne permet pas d'appliquer l'algorithme sur des variables catégoriques.

Ainsi nous décidons d'appliquer la transformation en Analyse en Composantes Multiples afin d'obtenir deux facteurs à valeurs réelles qui seront ensuite utilisés pour appliquer l'algorithme des K-Moyennes :

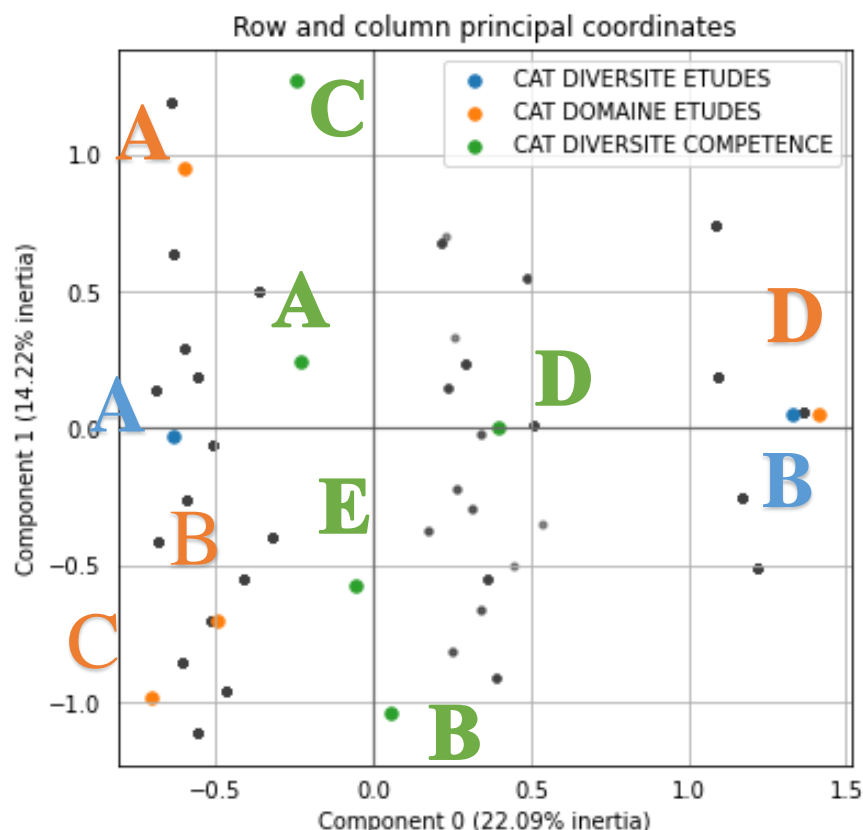


Figure 5-10. Graphique de l'Analyse en Composantes Multiples (ACM) à 2 facteurs sur les 3 variables catégoriques du Modèle 3

Pour la variable « Diversité d'études », la valeur A indique une faible diversité et la valeur B une forte diversité. Pour la variable « Diversité des compétences », la diversité des compétences augmente en suivant l'ordre alphabétique des catégories de A vers E. Enfin pour la variable « Domaine d'études », les catégories A, B et C désignent un seul domaine d'études respectivement en Aéronautique & Mécanique, Électrique et Informatique & Logiciel, et la catégorie D désigne plusieurs de ces domaines ou le domaine Business – Management - Finance.

On retrouve sur la figure les valeurs des catégories des 3 variables Diversité des études, Domaine d'études, Diversité des compétences dans le plan des 2 facteurs qui résultent de l'ACM.

Ensuite, nous appliquons alors l'algorithme des K-Moyennes en tant que méthode exploratoire pour former des classes sur notre ensemble d'entraînement. L'initialisation des centroïdes est faite aléatoirement sur l'ensemble d'entraînement. La fonction coût définie à la section 4.5.3, prend la forme suivante pour des expérimentations allant de K=1 classe à 50 classes.

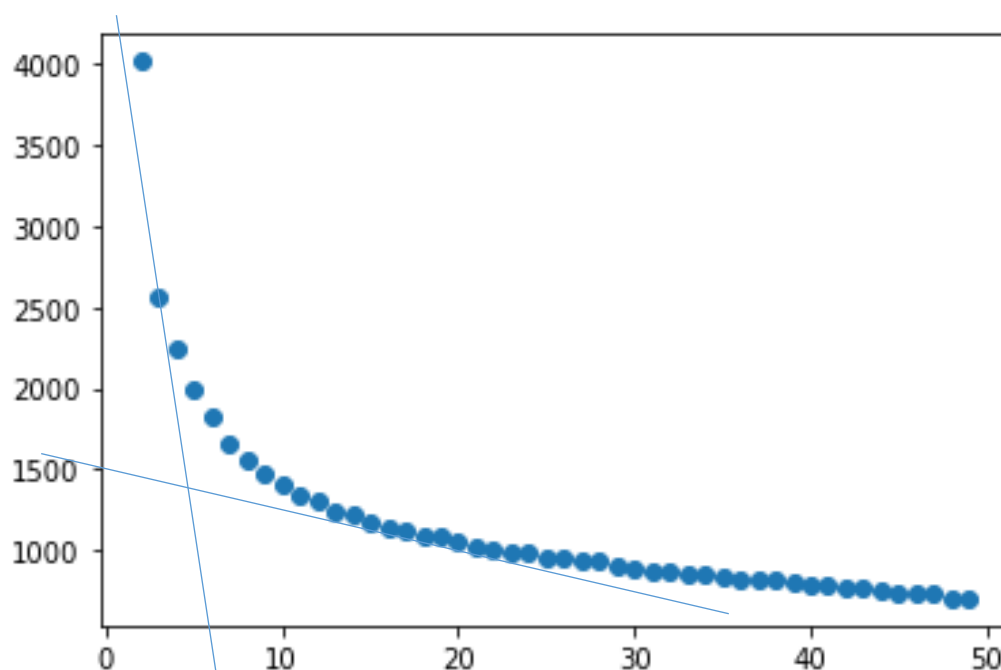


Figure 5-11. Graphe de la fonction coût de l'algorithme des K-Moyennes sur le Modèle 3.

Entropie en fonction du nombre de classes K.

On observe alors qu'un optimum du nombre de classes K serait autour de 5 classes. En effet, il s'agit de l'abscisse du point du graphe qui résulte de l'intersection des asymptotes (cf. Figure 5-10).

Ainsi, nous optons pour conserver cette valeur de classes $K=5$. A des fins d'analyses, nous observons les moyennes de nombre de réalisés pour chaque classe ainsi que la moyenne des valeurs des descripteurs du ratio entre le nombre de brevets dans le domaine favoris sur le nombre de brevets total pour chaque inventeur.

Tableau 5-2. Analyse des descripteurs du potentiel d'innovation pour chacune des classes issues de l'algorithme des K-Moyennes.

Classe	MOYENNE NOMBRE BREVETS	DESCRIPTEUR POTENTIEL D'INNOVATION du CO	Effectif (sur 85)
classe 4	7,7	0,412	17
classe 3	4,05	0,340	18
classe 2	2,33	0,360	15
classe 1	8,64	0,342	14
classe 0	6,81	0,772	21

Le tableau ci-dessus nous présente les différentes valeurs atteintes de la moyenne du nombre de brevets atteints par chaque classe, ainsi que la valeur du descripteur du potentiel d'innovation du CO qui prend une valeur réelle entre 0 et 1.

L'analyse que nous pouvons fournir de ces résultats est la suivante :

- Classe 4 : Moyenne de réalisation de brevets élevée & moyenne domaine favoris normal
=> profils très innovants et domaines plutôt spécialisés
- Classe 3 : Moyenne de réalisation de brevets normale & moyenne domaine favoris faible
=> profil innovant dans des domaines technologique complexes (peu de brevets sur ce sujet)
- Classe 2 : Moyenne de réalisation de brevets faible & moyenne domaine favoris faible => profils peu innovants
- Classe 1 : Moyenne de réalisation de brevets élevée & moyenne domaine favoris faible => profils très innovants et polyvalents
- Classe 0 : Moyenne de réalisation de brevets élevée & moyenne domaine favoris élevée => profils très innovants et très spécifiques

Afin de confirmer ces hypothèses établies à partir des indicateurs de publications brevets et descripteur du potentiel d'innovation du CO de l'entreprise par les publications brevets, nous analysons également les statistiques descriptives des variables catégoriques du modèle 3.

La figure ci-dessous représente la distribution des profils par classe selon leurs domaines d'études. Ainsi les couleurs désignent les catégories de la variable catégorique des domaines d'études.

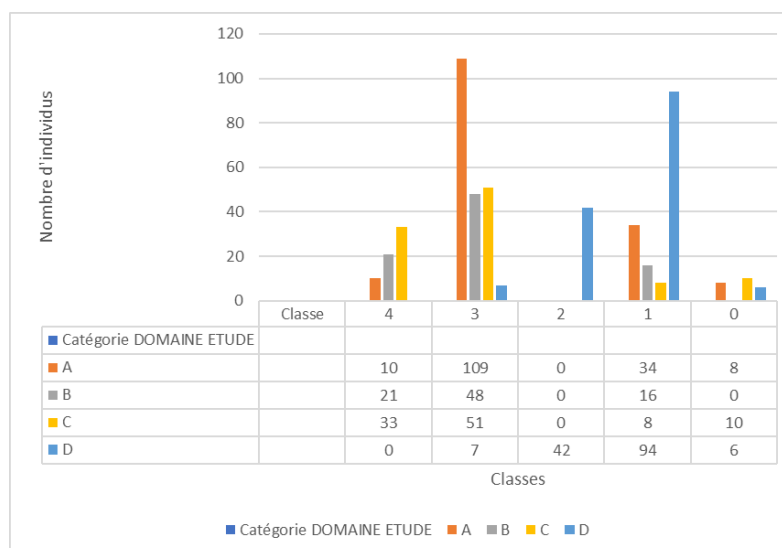


Figure 5-12. Distribution des profils par classe selon leurs domaines d'études

Ainsi par exemple, les employés de la classe 3 (hypothèse : profils innovations dans des domaines technologiques complexes) sont des personnes issues d'une formation en Aéronautique & Mécanique principalement, tandis que les employés de la classe 1 (hypothèse : profils très innovants et polyvalents) sont des employés issus d'une formation combinées de plusieurs domaines d'études. De plus, les individus de la classe 0 (hypothèse : profils très innovants et très spécifiques) auraient des profils ayant étudié dans une formation Aéronautique & Mécanique, Informatique & Logiciels et des combinaisons de plusieurs domaines techniques. Enfin les individus de la classe 2 ont tous étudiés dans plusieurs domaines ou en Business – Management – Finance, et les employés de la classe 4 ont tous étudié uniquement soit en Génie Aéronautique & Mécanique, Génie Électrique ou Génie Informatique & Logiciel.

Ce graphique a donc uniquement une valeur explicative en tant que statistiques descriptives, elles permettent d'évaluer d'où proviennent les profils de chaque classe. Nous allons voir que les statistiques descriptives suivantes fournissent des informations qui permettent de valider nos hypothèses.

La figure suivante présente la distribution des profils par classe selon la diversité des groupes de compétences. Il s'agit d'analyser si les employés d'une même classe ont tendance à posséder un grand panel de compétences ou bien des compétences dans un sujet plus spécifiques.

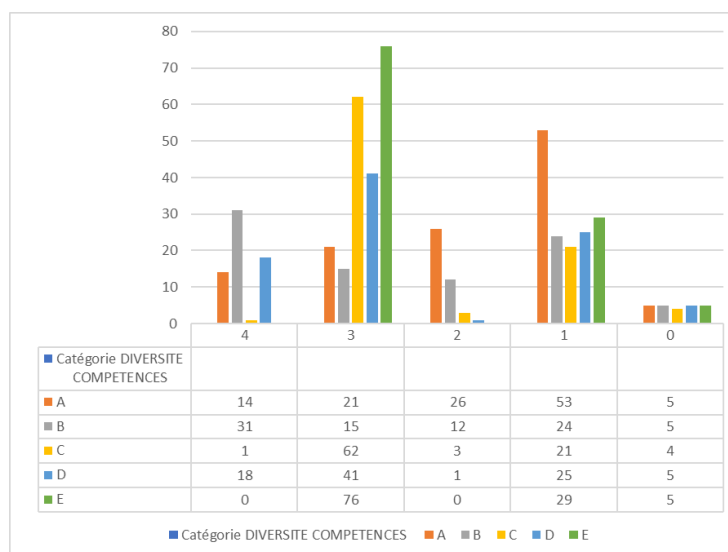


Figure 5-13. Distribution des profils par classe selon la Diversité des groupes de compétences

Pour rappel, la variable catégorique de Diversité des groupes de compétences a des catégories de diversité de plus en plus forte de la lettre A vers la lettre E. Ainsi pour la classe 3 (hypothèse : profils innovations dans des domaines technologiques complexes), les employés ont tendances à posséder une grande variété de groupes de compétences, tandis que pour la classe 1 (hypothèse : profils très innovants et polyvalents) les employés ont tout type de variétés de compétences. Quant à la classe 2 (hypothèse : profils peu innovants) et à la classe 4 (hypothèse : profils très innovants et domaines plutôt spécialisés), les individus ont peu voire très peu de variété de compétences. Enfin, la classe 0 présente des employés qui ont tout type de diversité de groupes de compétences et notamment un bon nombre d'individus qui ont une forte diversité de compétences (hypothèse : profils très innovants et très spécifiques).

Les statistiques descriptives de la variable catégorique « Diversité des groupes de compétences » est favorable pour la validation des hypothèses, car les classes d'individus aux profils innovants avec des domaines technologiques complexes et polyvalents ont majoritairement une forte diversité de groupes de compétences, contrairement aux profils plutôt spécialisés ou très spécifiques.

La figure suivante présente la distribution des profils par classe et selon le niveau d'études. Il s'agit ici d'analyser le niveau d'études au sein d'une même classe.

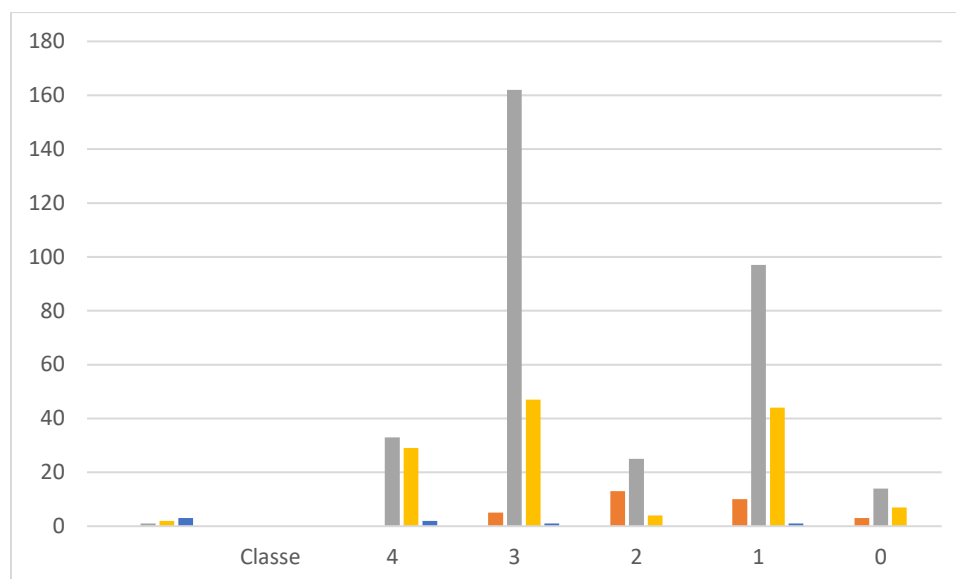


Figure 5-14. Distribution des profils par classe selon le niveau d'études

Proportionnellement, la classe 1 (hypothèse : profils très innovants et polyvalents), la classe 4 (hypothèse : profils très innovants et domaines plutôt spécialisés) et la classe 0 (hypothèse : profils très innovants et très spécifiques) possèdent beaucoup d'éléments qui ont atteint un grade de Master (barre jaune), tandis que la classe 3 (hypothèse : profils innovants dans des domaines technologiques complexes) a une grande majorité d'employés qui ont le grade de Baccalauréat. La classe 2 (hypothèse : profils peu innovants) quant à elle aurait tendance à rassembler des profils moins diplômés.

Ainsi, compte tenu des résultats pour la variable discrète du niveau d'études les hypothèses formulées sur les tendances d'innovations semblent suivre une certaine logique, plus le grade atteint par l'individu est élevé plus l'individu aurait un potentiel d'innovation fort.

La figure qui suit propose la distribution des profils par classe selon la diversité des études. Il s'agit d'observer sur les employés d'une même classe ont tendance à avoir étudié dans différents domaines d'études ou à se concentrer sur un domaine particulier.

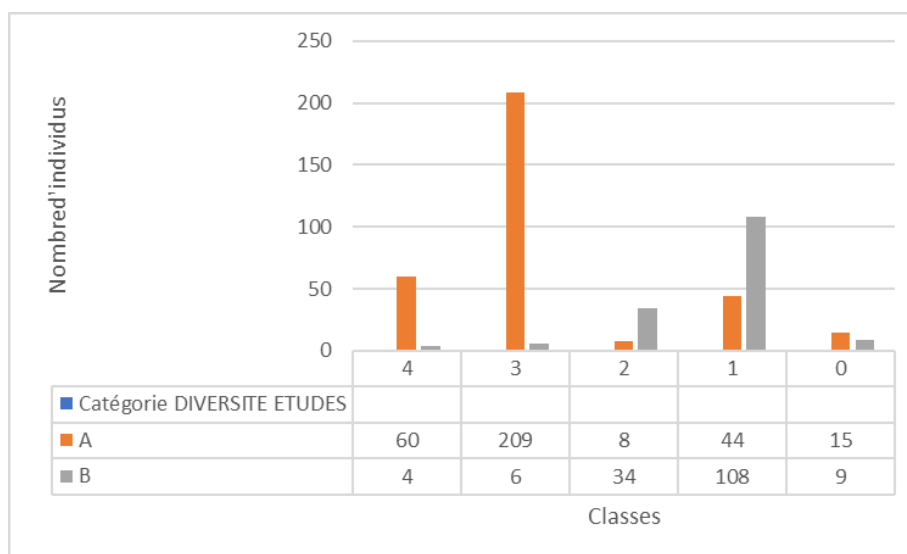


Figure 5-15. Distribution des profils par classe selon la diversité des études

Nous constatons sur cette distribution que les employés de la classe 4 (hypothèse : profils très innovants et domaines plutôt spécialisés) et la classe 3 (hypothèse : profils innovants dans des domaines technologiques complexes) ont une formation très majoritairement spécifique tandis que les classes 1 (hypothèse : profils très innovants et polyvalents) et 2 (hypothèse : profils peu innovants) comportent des individus qui ont réalisé des formations davantage diversifiées avec plusieurs sujets d'études. Enfin la classe 0 (hypothèse : profils très innovants et très spécifiques) comporte des employés qui auraient effectué qu'un seul sujet d'études.

Finalement, la variable catégorique « Diversité des études » nous apportent également des éléments favorables pour valider les hypothèses formulées à savoir que les profils qui sont innovants sur des sujets spécifiques ou plutôt spécialisés, auraient étudié un sujet d'études unique tandis que les profils innovants polyvalents et complexes auraient des sujets d'études diversifiés.

CHAPITRE 6 DISCUSSION, CONCLUSION ET RECOMMANDATIONS

Ce chapitre est réservé à l'interprétation des résultats, à une discussion sur la validité et la fiabilité de la recherche ainsi qu'à des recommandations et la description des limites de la recherche.

6.1 Interprétation des résultats

Les résultats obtenus à l'aide des règles d'associations permettent de vérifier certaines règles basiques sans apporter des informations non-intuitives. En effet, les résultats qui associent les études en Informatique et Logiciel vers la publication de brevets ne sont pas d'une grande aide afin de déterminer sur un ensemble d'individus leur potentiel d'innovation. La méthode des règles d'associations pourrait s'avérer utile dans le cadre d'un ensemble de données plus important. On pourrait imaginer également d'autres indicateurs d'innovation plus précis à l'échelle d'un projet de l'entreprise, qui ne se limitent pas à la dimension de publication de brevets pour le potentiel d'innovation du CO d'une entreprise afin d'obtenir des règles d'association ou des classes (selon la méthode d'apprentissage utilisée) qui pourraient avoir des cas d'usage pratiques à l'entreprise industrielle telle que la constitution d'équipe pour des projets particuliers.

De même, les résultats obtenus avec le modèle 2 par les arbres de décisions ne sont pas suffisamment discriminants afin de tirer des éléments prédictifs pour le potentiel d'innovation.

D'après les analyses exploratoires menées, seulement le troisième modèle qui fait recours à des variables binaires, discrètes et catégoriques pour les variables du CH et pour les indicateurs de publications de brevets pour le potentiel d'innovation du CO amène à des interprétations plus intéressantes. La méthode d'apprentissage, nous amène à retenir un modèle de classification en 5 classes de notre ensemble d'entraînement.

En effet, les résultats que nous avons pu obtenir à l'issue de l'algorithme des K-Moyennes (décrits à la section 5.2.3) proposent davantage d'information que les méthodes précédentes. Afin de savoir si ce modèle d'apprentissage peut être utilisé à des fins prédictives, il convient d'évaluer les hypothèses que nous avons formulées. Pour ce faire nous proposons d'évaluer les hypothèses grâce aux statistiques descriptives des valeurs des variables du CH pour chacune des 5 classes.

- Classe 4 (hypothèse : profils très innovants et domaines plutôt spécialisés)

Les domaines d'études de ces personnes sont concentrés dans un seul domaine du génie car seules les catégories A (Aéronautique & Mécanique), B (Électrique) et C (Informatique & Logiciel) sont présentes, d'après les statistiques descriptives de la variable catégorique « Domaines d'études ». Il y a une majorité de profils issus d'une formation en Génie Informatique & Logiciels, d'après la majorité de faible « Domaine d'études ».

De plus, ces profils ont relativement des compétences très ciblées et peu diversifiées, majorité de catégorie B, absence de catégorie E (maximum de diversité de compétences) respectivement à la construction de la variable catégorique « Diversité des groupes de compétences ».

> **Les 3 variables catégoriques représentent bien cette hypothèse**

- Classe 3 (hypothèse : profils innovants dans des domaines technologiques complexes)

Pour les domaines d'études, on observe une distribution similaire à la classe 4 avec une majorité de domaine d'études dans un seul génie. Toutefois on observe une majorité de profils issus d'une formation en génie aéronautique ou mécanique, catégorie A.

On observe également une grande majorité de catégorie A en diversité des études, ce qui signifie que ces profils se concentrent dans un domaine d'études particulier.

De plus, ces profils ont des compétences nombreuses et diverses, à la fois dans plusieurs domaines techniques, notamment dans le secteur aéronautique, en logiciels d'ingénierie, ainsi qu'en management et relationnel. Il y a une majorité de catégories diversité de compétences C, D et E (sachant que la diversité des compétences variant de A – très faible à E – très forte).

> **Les 3 variables catégoriques représentent bien cette hypothèse**

- Classe 2 (hypothèse : profils peu innovants)

On remarque seulement des profils ayant obtenu des diplômes dans plusieurs domaines de génie combiné à un diplôme en Business-Management-Finance, catégorie D.

En conséquence, pour la « Diversité des études » on retrouve une majorité de catégorie B (forte diversité des domaines d'études). De plus, ces profils ont relativement peu de compétences car il y

a une majorité de catégorie A (pas de catégorie de compétence) et très faible proportion de catégories B, C et D.

- > **La variable catégorique « Diversité des compétences » peut justifier pourquoi ces profils ne seraient pas innovants**

- Classe 1 (hypothèse : profils très innovants et polyvalents)

On retient une grande majorité de profils ayant obtenus plusieurs diplômes en domaine de génie et Business, catégorie D ; avec quelques profils issus des domaines Aéronautique & Mécanique, Électrique, Informatique & Logiciel.

De plus, la majorité de ces profils ont une forte diversité de leurs domaines d'études (ils étudient plusieurs domaines d'études). Pour les compétences, on observe une distribution homogène des catégories de diversité, ce qui signifie que ces profils ont une tendance à acquérir une multitude de compétences à la fois techniques, managériales et relationnelles.

- > **Les 3 variables catégoriques représentent bien cette hypothèse**

- Classe 0 (hypothèse : profils très innovants et très spécialisés – experts de leur domaine)

La distribution de la variable « Domaine d'études » est homogène dans les catégories A, C et D ; ce sont profils qualifiés en génie aéronautique-mécanique, génie informatique ou des combinaisons de diplômes en génie et Business. On remarque une absence de profils ayant seulement un diplôme en génie électrique. De plus, on observe une distribution homogène sur la diversité des études. Pour les compétences, ces profils disposent de compétences diverses d'après la distribution homogène sur la variable « Diversité des compétences ».

- > **Analyser d'autres variables pour interpréter les caractéristiques de la classe 0. Les 3 variables catégoriques ne semblent pas suffisantes pour expliquer pourquoi ces individus peuvent être considérés comme expert de leur domaine**

Le tableau ci-dessous reprend l'ensemble des hypothèses formulés et les résultats obtenus avec les analyses menées avec l'algorithme de classification des K-Moyennes sur le Modèle 3.

Tableau 6-1. Hypothèses et résultats obtenus avec l'algorithme de classification sur le Modèle 3

Hypothèse	Résultats
Hypothèse 1: La classe 4 présente des profils très innovants et domaines plutôt spécialisés	Les employés de la classe 4 ont étudié dans un seul domaine d'études provenant du domaine Informatique & Logiciel principalement, ils ont des compétences très ciblées et peu diversifiées
Hypothèse 2 : La classe 3 intègre des profils innovants dans des domaines technologiques complexes	Les individus de la classe 3 ont majoritairement étudié dans un seul domaine d'études provenant du domaine Aéronautique & Mécaniques. Ces personnes se sont concentrées sur un seul domaine d'étude (peu diversifié). Ces profils ont des compétences nombreuses et diverses
Hypothèse 3 : La classe 2 possède des profils peu innovants	Les individus de la classe 2 ont des diplômes dans plusieurs domaines, ils ont donc une grande diversité des domaines d'études et ces personnes sont qualifiés dans très peu de groupe de compétences.
Hypothèse 4 : La classe 1 inclue des profils très innovants et polyvalents	Les individus de la classe 1 possède en grande majorité des diplômes à la fois dans un domaine technique et en Business-Management-Finance. De plus, ils ont une forte diversité des domaines d'études. On retrouve également une distribution homogène sur les groupes de compétences, ce qui signifie que ces personnes acquièrent des compétences techniques, managériales et relationnelles.
Hypothèse 5 : La classe 0 comporte des profils très innovants et très spécialisés – experts de leur domaine	Le faible nombre d'individus dans la classe 0 et les statistiques descriptives analysées sur les variables catégoriques ne semblent pas discriminantes pour valider cette hypothèse

Les résultats nous indiquent que nous pouvons valider les hypothèses 1 à 4. Toutefois les statistiques descriptives établies en fonction des variables catégoriques (« Diversité des compétences, « Domaines d'études », « Diversité des domaines d'études ») ne permettent pas de vérifier que l'hypothèse 5 est vérifiée.

6.2 Discussion

6.2.1 Évaluation de la qualité et de la validité de l'étude

Validité interne

La collecte de données du CH de LinkedIn et des données brevets de Patseer sont effectuées à un instant bien précis. Or les dimensions que nous souhaitons mesurer sont des éléments qui évoluent au cours du temps. En effet, les compétences d'une personne, ses formations, ses expériences professionnelles et même ses publications sont tous des éléments qui sont en perpétuels changements. Ainsi le modèle utilisé est soumis à un **effet de maturation**, également appelée effet de temps. Il s'agit d'une erreur survenant lorsque la manipulation de la variable indépendante (données du CH) n'entraîne pas la modification de la variable dépendante (potentiel d'innovation du CO) dans un laps de temps mesuré pendant l'enquête (qui peut être de plusieurs dizaines d'années dans notre cas). En effet, il se peut que les inventeurs de brevets de l'entreprise industrielle étudiée ne soient plus en activité professionnelle au moment de l'étude. Ainsi, les dimensions du CH (compétences, parcours, etc.) qui ont eu un impact pour la publication du brevet pourraient s'avérer erronées ou bien même ne plus exister au moment de l'étude. Ce qui signifierait que pour la publication de brevet dans le même domaine technologique au moment de l'analyse, posséder les mêmes valeurs pour les dimensions du CH qu'un profil innovant à l'époque ne serait pas suffisant pour être suffisamment innovant afin de publier un brevet dans ce domaine technologique aujourd'hui. Ceci s'explique par l'évolution très rapide des écosystèmes industriels au cœur de l'innovation technologique.

Fiabilité globale de la recherche

La fiabilité globale de la recherche suppose que la collecte des données de LinkedIn et des brevets de Patseer soit entièrement reproductible et aux mesures des différentes variables définies.

La méthodologie adoptée pour la collecte et le traitement des données a été entièrement décrit dans le chapitre 4. Ce chapitre montre toutes les étapes de sélection, de traitement des données, de création des variables ainsi que les méthodes d'analyse utilisées. Enfin, les résultats d'analyse montrent le sens statistique et le niveau de signification de chacun des coefficients.

6.2.2 Recommandations

Les recommandations à émettre pour les études futures sur le sujet de recherche d'identification du potentiel d'innovation du CO de l'entreprise à partir du CH de ses employés concernent notamment le choix des variables pour caractériser le concept du CH. Comme nous l'avons vu dans la revue de littérature sur le CI des entreprises et des études sur le lien entre le CH et le CO de l'entreprise dans le cadre du processus d'innovation. Le CH est un concept qui est difficile à mesurer de par la complexité de la nature humaine et des indicateurs de mesure à disposition pour mesurer des dimensions intangibles du CH.

Plus précisément, à propos des variables il serait possible de construire des variables qui tiennent compte des expériences professionnelles de l'employé, de par son poste et également son expérience. De nombreuses informations sont disponibles sur les profils LinkedIn et n'ont pas été prises en compte dans cette première étude sur le sujet. Les rubriques visibles sur les profils LinkedIn peuvent entre autres apporter des informations sur les éventuels projets, publications et brevets réalisés par l'employé. Il serait également possible d'envisager des méthodes de traitement de texte pour analyser le contenu des résumés des profils des employés ou encore d'analyser avec précision le détail des descriptions, des expériences professionnelles, les descriptions des formations, le contenu de certains projets ou encore des associations. Ce sont des données qui viendraient compléter le jeu de données utilisé lors de ces études qui recouvrent seulement une partie de l'information disponible sur un profil.

Évidemment, il serait bénéfique de bénéficier d'un plus grand nombre de données afin d'avoir un modèle analyse plus robuste et plus représentatif du contenu des données que l'on trouve sur LinkedIn ou sur les bases de données brevets. Toutefois, ceci nécessiterait davantage de moyens pour accéder aux différentes sources de données à grande échelle.

D'autre part, il serait intéressant de compléter les variables du concept de potentiel d'innovation du CO de l'entreprise. En effet, lors de cette étude exploratoire le concept de potentiel d'innovation du CO est envisagé uniquement sous la dimension des publications de brevets, toutefois il existe de nombreuses autres dimensions qui pourraient servir d'indicateur au potentiel d'innovation en entreprise industrielle. Ceci représente d'autres pistes à envisager dans les études exploratoires au sujet de lier le CH des employés au potentiel d'innovation de l'entreprise.

6.3 Conclusion et limites de la recherche

6.3.1 Conclusion

Le mémoire propose une modélisation afin d'identifier le potentiel d'innovation du CO de l'entreprise industrielle selon 5 types de profils, à partir des données du CH des employés. Cette modélisation est à la fois conditionnée par la publication de brevets de l'entreprise et la source de données brevets Patseer et par la collecte et le traitement de données du CH provenant de LinkedIn qui s'effectue par l'usage du Moissonneur Web (qu'il faut maintenir à jour afin d'assurer la disponibilité des données du CH). Les résultats obtenus permettent de découvrir des tendances d'innovation des employés générales en fonction des dimensions de Développement (formation) et du Savoir-faire (compétences) du concept de CH. Cette étude exploratoire qui est la seule dans la littérature à invoquer le concept d'innovation en lien avec le CH à l'échelle individuelles (des employés) d'une entreprise suggère donc qu'il est possible d'extraire des connaissances à partir de ces sources de données. Toutefois, une entreprise qui souhaite obtenir un outil d'aide à la décision en vue d'appliquer une stratégie d'innovation devrait approfondir ces pistes de recherche. En effet, la modélisation obtenue en l'état n'apporte pas de précision particulière aux projets de l'entreprise car le modèle de notre étude exploratoire n'implique pas de variable à ce sujet.

6.3.2 Limites de la recherche

La méthodologie de l'étude exploratoire présente plusieurs limites qu'il semble important de mentionner. Tout d'abord, la collecte des données de l'ensemble des profils LinkedIn des employés de l'entreprise analysée est soumise au respect des lois locales et de l'éthique mise en place par le réseau professionnel LinkedIn. L'outil d'extraction de données sur le Web, dit *Moissonneur*, qui a été mis en place pour collecter les données sur le réseau professionnel respecte la confidentialité des données personnelles des utilisateurs et n'entrave pas la protection des données que met en place le site LinkedIn pour les utilisateurs qui souhaitent conserver leurs données privées. Ceci constitue une limite au modèle proposé car il s'agit d'un biais de non-divulgaration. Certains utilisateurs de LinkedIn protègent leurs données du public et par conséquent, nous ne pouvons collecter les informations de leurs profils afin de compter ces utilisateurs dans l'ensemble pour mener notre étude exploratoire. L'ensemble est donc limité aux profils LinkedIn des employés de l'entreprise qui rendent leurs données publiques sur le réseau professionnel. Il s'agit d'un premier

aspect du biais de non-divulgateion. De plus, à cet égard, les données publiques collectées sur le réseau professionnel sont anonymisées afin de respecter la confidentialité de ces personnes. Le nombre de données que nous pouvons utiliser pour notre modèle d'apprentissage est donc limité aux données des employés de l'entreprise industrielle qui accepte l'utilisation de leurs données.

Par ailleurs, le contenu des données du CH brut provenant de la source LinkedIn présente un inconvénient qui réside dans le fait que les utilisateurs peuvent ne pas dévoiler des informations correctes à leur égard ou à l'égard d'autres utilisateurs. Il s'agit du biais de non-divulgateion ou désirabilité sociale rendent l'information de nature particulièrement subjective ce qui peut être une limite importante en recherche alors que l'on vise à atteindre une certaine forme de vérité objective. La non-divulgateion d'éléments négatifs pertinents ou d'éléments positifs considérés comme non pertinents peut affecter le potentiel de cette source de données, car elle limite le nombre de données contextuelles qu'il est possible d'obtenir. Par contre, le biais de désirabilité sociale induit un autre enjeu plus important encore, soit l'intention de vouloir se signaler de la façon la plus positive que possible, quitte à induire une nouvelle réalité si celle-ci ne reflète pas ce qui est perçu d'être attendu de l'organisation. Ceci signifie qu'une information véhiculée sur un site internet peut être tout à fait vraie, partiellement vraie, projetée (pas encore vraie, mais a le potentiel pour le devenir) ou tout simplement fausse.

RÉFÉRENCES

- Abuzyarova, D., Belousova, V., Krayushkina, Zh., Lonsheikova, Y., Nikiforova, E., & Chichkanov, N. (2019). The Role of Human Capital in Science, Technology and Innovation. *Foresight and STI Governance*, 13(2), 107–119.
- Agrawal, R., & Srikant, R. (2000). Fast Algorithms for Mining Association Rules. *Proc. 20th Int. Conf. Very Large Data Bases VLDB*, 1215.
- Ahuja, G. (2000). Collaboration Networks, Structural Holes, and Innovation: A Longitudinal Study. *Administrative Science Quarterly*, 45(3), 425–455. <https://doi.org/10.2307/2667105>
- Alegre, J., & Lapiedra, R. (2005). Gestión del conocimiento y desempeño innovador: un estudio del papel mediador del repertorio de competencias distintivas. *Cuadernos de Economía y Dirección de la Empresa*, 23, 117-138.
- Anderson, R. (2008). Implications of the information and knowledge society for education. In J. Voogt, & G. Knezek (Eds.), *International handbook of information technology in primary and secondary education* (pp. 5e22). New York: Springer.
- Arifovic, J. (1994). Genetic algorithm learning and the cobweb model. *Journal of Economic dynamics and Control*, 18(1), 3-28.
- Becker, G. S. (1964). Human capital: A theoretical and empirical analysis, with special reference to education. *New York: National Bureau of Economic Research*.
- Berry, M.J.A., & Linoff, G. (2004). Data Mining Techniques: For Marketing, Sales, and Customer Support, *New York: Wiley*.
- Bozbura, F. T. (2004). Measurement and application of intellectual capital in Turkey. *The Learning Organization*, 11(4/5), 357-367.
- Bontis, N. (1998). Intellectual capital: an exploratory study that develops measures and models. *Management decision*, 36(2), 63-76.
- Bueno, E., Ordóñez de Pablos, P., & Salmador, M.P. (2004). Towards an integrative model of business, knowledge and organizational learning processes. *International Journal of Technology Management*, 27 (6/7), 562-574.

- Brin, S., Motwani, R., Ullman, J. D., & Tsur, S. (1997). Dynamic itemset counting and implication rules for market basket data. *ACM SIGMOD Record*, 26(2), 255–264. <https://doi.org/10.1145/253262.253325>
- Brunet, F., & Malas, K. (édit.). (2019). *L'innovation en santé : Réfléchir, agir et valoriser*. (2^{ème} éd.). CHU Sainte-Justine
- Brynjolfsson, E., & McAfee, A. (2014). The second machine age: Work, progress, and prosperity in a time of brilliant technologies. *WW Norton & Company*.
- Camisón, C., Palacios, D., Devece, C., (2000). Un nuevo modelo para la medición del Capital Intelectual en la empresa: El modelo Nova. *X Congreso Nacional ACEDE*. Oviedo, September.
- Cardinal, L. B. (2001). Technological innovation in the pharmaceutical industry: The use of organizational control in managing research and development. *Organization Science*, 12 : 19–36.
- Cheeseman, P., Kelly, J., Self, M., Stutz, J., Taylor, W., & Freeman, D. (1988). Autoclass: A Bayesian classification system. In *Machine learning proceedings 1988* (pp. 54-64). Morgan Kaufmann.
- Chelmiss, C., Sorathia, V., & Prasanna, V. K. (2012) “Enterprise wisdom captured socially,” in *Proc. IEEE/ACM Int. Conf. Adv. Soc. Netw. Anal.Min.*, Istanbul, Turkey, 1228–1235.
- Claver-Cortes, E., Zaragoza-Saez, P. C., Molina-Manchon, H., & Ubeda-Garcia, M. (2015). Intellectual capital in family firms: human capital identification and measurement. *Journal of Intellectual Capital*, 16(1), 119-223.
- Cohendet, P., Grandadam, D., & Simon, L. (2010). The Anatomy of the Creative City, *Industry and Innovation*, 17:1, 91-111, DOI: 10.1080/13662710903573869
- Cohendet, P., & Simon, L. (2020). *New Spaces of Innovation*, MOSAIC 2020 : HEC Montréal
- Delaplace, M. (2008). Innovations organisationnelles et radicales : Le cas de deux grandes firmes de l'informatique émergente. *Revue française de gestion*, 185-200. <https://doi.org/10.3166/rfg.188-189.185-200>
- Diday, E. (1971). Une nouvelle méthode en classification automatique et reconnaissance des formes la méthode des nuées dynamiques. *Revue de statistique appliquée*, 19(2), 19-33.

- Dougherty D. (1992). Interpretive Barriers to Successful Product Innovation in Large Firms. *Organization Science* 3 (2), 179-202 <https://doi.org/10.1287/orsc.3.2.179>
- Euroforum (1998), Medición del Capital Intelectual, *Instituto Universitario Euroforum El Escorial*, Madrid.
- Fages, R., Terrenoire, M., Tounissoux, D., & Zighed, A. (1987). Non supervised classification tools adapted to supervised classification. In *Pattern Recognition Theory and Applications* (pp. 57-62). Springer, Berlin, Heidelberg.
- Gamerschlag, R. (2013). Value relevance of human capital information. *Journal of Intellectual Capital*, 14(2), 325-345.
- Georgescu, R. M., Savu, I. C., & Militaru, G. (2018) How Can Social Networks Improve The Recruitment Process Case Study - Linkedin, *Proceedings of the International Management Conference*, Faculty of Management, Academy of Economic Studies, Bucharest, Romania, 12(1), 644-652, November.
- Giddings, F. H. (1890). Principles of Economics. By Alfred Marshall, Professor of Political Economy in the University of Cambridge. MacMillan & Co., London and New York, 1890. Vol. I, pp. xxviii, 754. *The Annals of the American Academy of Political and Social Science*, 1(2), 332–337. <https://doi.org/10.1177/000271629000100217>
- Ehrlich, K. Lin, C.-Y. & Griffiths-Fisher, V. (2007). “Searching for experts in the enterprise: Combining text and social network analysis”, in *Proc. ACM Conf. Supporting Group Work*, Sanibel Island, FL, 117–126.
- Helfat, C. E., & Martin, J. A. (2015). Dynamic Managerial Capabilities: Review and Assessment of Managerial Impact on Strategic Change. *Journal of Management*, 41(5), 1281–1312. <https://doi.org/10.1177/0149206314561301>
- Issor, Z. (2017). La performance de l’entreprise: un concept complexe aux multiples dimensions. *Projectics/Proy ctica/Projectique*, (2), 93-103.
- Kaplan, R. S., & Norton, D. P. (2001). Transforming the balanced scorecard from performance measurement to strategic management: Part 1. *Accounting horizons*, 15(1), 87-104.

- Kanagawa, Y., Matsumoto, S., Koike, S., & Imamura, T. (2009). Association analysis of food allergens. *Pediatric Allergy and Immunology*, 20(4), 347–352. <https://doi.org/10.1111/j.1399-3038.2008.00791.x>
- Karnouskas S. (2017). Massive open online courses (MOOCs) as an enabler for competent employees and innovation in industry. *Computers in Industry*, 91, 1–10.
- Le Chapelain, C. & Matéos, S. (2020). Schultz et le capital humain : une trajectoire intellectuelle. *Revue d'économie politique*, 130, 5-25. <https://doi.org/10.3917/redp.301.0005>
- Lev, B. (2001). *Intangibles : Management, Measurement, and Reporting*. Washington, D.C.: Brookings Institution Press. Retrieved March 28, 2021, from <http://www.jstor.org/stable/10.7864/j.ctvcj2rf2>
- Levitt, T. (1976). Alfred Marshall: Victorian Relevance for Modern Economics. *The Quarterly Journal of Economics*, 90(3), 425–443. <https://doi.org/10.2307/1886042>
- Li, L. (2018). China's manufacturing locus in 2025: With a comparison of “Made-in-China 2025” and “Industry 4.0”. *Technological Forecasting and Social Change*, 135, 66–74. <https://doi.org/10.1016/j.techfore.2017.05.028>
- Lin, C.-Y., Wu, L., Wen, Z., Tong, H., Griffiths-Fisher, V., Shi, L., & Lubensky, D. (2012) “Social network analysis in enterprise,” *Proc. IEEE*, vol. 100(9), 2759–2776
- Madhavan, R., & Grover, R. (1998). From Embedded Knowledge to Embodied Knowledge: New Product Development as Knowledge Management. *Journal of Marketing*, 62(4), 1-12. <https://doi.org/10.2307/1252283>
- Mariz-Perez, R. M., Teijeiro-Alvarez, M. M., García-Alvarez, & M. Teresa. (2012). The relevance of human capital as a driver for innovation. *Cuadernos De Economía*, 35(98), 68–76. [https://doi.org/10.1016/S0210-0266\(12\)70024-9](https://doi.org/10.1016/S0210-0266(12)70024-9)
- Massingham, P., Nguyet Que Nguyen, T., & Massingham, R. (2011). Using 360 degree peer review to validate self-reporting in human capital measurement. *Journal of Intellectual Capital*, 12(1), 43-74.
- Michalski, R. S. (1983). A theory and methodology of inductive learning. In *Machine learning* (pp. 83-134). Morgan Kaufmann.

- Mount, M. K., Barrick, M. R., & Stewart, G. L. (1998). Five-factor model of personality and performance in jobs involving interpersonal interactions. *Human performance*, 11(2-3), 145-165.
- Nahapiet, J., & Ghoshal, S. (1998). Social capital, intellectual capital, and the organizational advantage. *Academy of Management Review*, 23: 242–266.
- Nonaka, I., & von Krogh, G. (2009). Tacit knowledge and knowledge conversion: controversy and advancement in organizational knowledge creation theory. *Organization Science*, 20(3), 635–652.
- OCDE/Eurostat (2019). Manuel d’Oslo 2018 : Lignes directrices pour le recueil, la communication et l’utilisation des données sur l’innovation, 4ème édition, *Mesurer les activités scientifiques, technologiques et d'innovation*, Éditions OCDE, Paris, <https://doi.org/10.1787/c76f1c7b-fr>
- O*NET (2018). The O*NET® Content Model, <https://www.onetcenter.org/content.html> (consulté le 6 septembre 2021).
- Ostrovsky, R., & Rabani, Y. (2006). “The effectiveness of lloyd-type methods for the k-means problem,” in *47th IEEE Symposium on the Foundations of Computer Science (FOCS)*, pp. 165–176.
- Partnership for 21st century skills. (2008). 21st century skills, education & competitiveness: A resource and policy guide.
- Pisano, S., Lepore, L., & Lamboglia, R. (2017). "Corporate disclosure of human capital via LinkedIn and ownership structure: An empirical analysis of European companies", *Journal of Intellectual Capital*, Vol. 18 No. 1, pp. 102-127. <https://doi.org/10.1108/JIC-01-2016-0016>
- Roos, J., Edvinsson, L., & Dragonetti, N. C. (1997). *Intellectual capital: Navigating the new business landscape*. Springer.
- Simon, L. (2009). Underground, upperground et middle-ground : les collectifs créatifs et la capacité créative de la ville. *Management international / Gestión Internacional / International Management*, 13, 37–51. <https://doi.org/10.7202/037503ar>
- Subramaniam, M., & Venkatraman, N. (2001). Determinants of transnational new product development capability: testing the influence of transferring and deploying tacit overseas knowledge. *Strat. Mgmt. J.*, 22: 359-378. <https://doi.org/10.1002/smj.163>

Subramaniam, M., & Youndt, M. A. (2005). The Influence of Intellectual Capital on the Types of Innovative Capabilities. *Academy of Management Journal*, 48(3), 450–463. <https://doi.org/10.5465/amj.2005.17407911>

Suciu MC., & Năsulea DF. (2019). Intellectual Capital and Creative Economy as Key Drivers for Competitiveness Towards a Smart and Sustainable Development: Challenges and Opportunities for Cultural and Creative Communities. In: Matos F., Vairinhos V., Selig P., Edvinsson L. (eds) *Intellectual Capital Management as a Driver of Sustainability*. Springer, Cham. https://doi.org/10.1007/978-3-319-79051-0_5

Summa, L. (2016). Digitale Führungsintelligenz: “Adapt to win” – Wie Führungskräfte sich und ihr Unternehmen fit für die digitale Zukunft machen (S. 1). Wiesbaden: Springer.

Sveiby, K. E. (1997). The intangible assets monitor. *Journal of Human Resource Costing & Accounting*, 2(1).

Tsai, W., & Ghoshal, S. (1998). Social Capital and Value Creation: The Role of Intrafirm Networks. *The Academy of Management Journal*, 41(4), 464-476. <https://doi.org/10.2307/257085>

Tseng, C., & Goo, Y.J. (2005). Intellectual capital and corporate value in an emerging economy: empirical study of Taiwanese manufacturers. *R&D Management*, 35, 187-201

van Laar, E., Van Deursen, A. J. A. M., Van Dijk, J. A. G. M., & De Haan, J. (2017). The relation between 21st-century skills and digital skills: A systematic literature review. *Computers in Human Behavior*, 72, 577–588. <https://doi.org/10.1016/j.chb.2017.03.010>

Varshney, K. R., Wang, J., Mojsilović, A., Fang, D., & Bauer, J.H. (2013) “Predicting and recommending skills in the social enterprise,” in *Proc. Int. AAAI Conf. Weblogs Soc. Med.*, Cambridge, MA

Wang, J., Varshney, K. R., Mojsilović, A., Fang, D., Bauer, J. H. (2013). Expertise assessment with multi-cue semantic information, *Proceedings of 2013 IEEE International Conference on Service Operations and Logistics, and Informatics*, 534-539, doi: 10.1109/SOLI.2013.6611472.

WEF (2017). The Global Human Capital Report. Preparing People for the Future of Work, Cologni: World Economic Forum.

Xu, J. (2014). Managing Digital Enterprise: Ten Essential Topics. Australia: Atlantis Press.

Youndt, M. A., Subramaniam, M., & Snell, S. A. (2004). Intellectual capital profiles: An examination of investments and returns. *Journal of Management Studies*, 41, 335–362.