

Titre: Proposition d'une méthodologie pour le développement d'une
plateforme de partage d'information au sein d'un réseau
décentralisé de PME
Title:

Auteur: Aimé Toumelin
Author:

Date: 2022

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Toumelin, A. (2022). Proposition d'une méthodologie pour le développement
d'une plateforme de partage d'information au sein d'un réseau décentralisé de
PME [Mémoire de maîtrise, Polytechnique Montréal]. PolyPublie.
Citation: <https://publications.polymtl.ca/10520/>

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

URL de PolyPublie: <https://publications.polymtl.ca/10520/>
PolyPublie URL:

**Directeurs de
recherche:** Bruno Agard, & Maxime Leduc
Advisors:

Programme: Maîtrise recherche en génie industriel
Program:

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

**Proposition d'une méthodologie pour le développement d'une plateforme de
partage d'information au sein d'un réseau décentralisé de PME**

AIMÉ TOUMELIN

Département de mathématiques et de génie industriel

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

Août 2022

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Ce mémoire intitulé :

**Proposition d'une méthodologie pour le développement d'une plateforme de
partage d'information au sein d'un réseau décentralisé de PME**

présenté par **Aimé TOUMELIN**

en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
a été dûment accepté par le jury d'examen constitué de :

Martin TRÉPANIER, président

Bruno AGARD, membre et directeur de recherche

Maxime LEDUC, membre et codirecteur de recherche

Christophe DANJOU, membre

DÉDICACE

*Ce mémoire est dédié à toutes les personnes
qui m'ont soutenues au cours de mes études,
ma famille ainsi que mes amis*

REMERCIEMENTS

Tout d'abord, je remercie le Professeur Agard pour les conseils qui m'ont été promulgués au cours de ce projet de recherche ainsi que pour le cadre exceptionnel de travail qu'il m'a fourni. Je suis aussi très reconnaissant pour l'atmosphère extraordinaire qu'il a su mettre en place dans son laboratoire. Je le remercie notamment pour ses expertises industrielles et scientifiques qui ont su guider ce projet de recherche. Ensuite, je remercie M. Leduc ainsi que l'entreprise Jasons systèmes fourrager pour avoir initié ce projet de recherche. Sa détermination, ainsi que son leadership tout au long de ce projet ont su transmettre une vision claire du projet permettant de le concrétiser. Je suis notamment très reconnaissant pour la passion qu'il a su initier dans ce projet ainsi que l'expertise agronomique qu'il m'a transmise. Je remercie le comité scientifique ainsi que le comité d'expert agronomique qui ont suivi l'évolution de ce projet et ont su me transmettre leurs expertises scientifiques. Je remercie l'entreprise Logiag pour avoir financé et fourni les outils technologiques nécessaires au développement de ce projet. Enfin, je remercie l'organisme MITACS pour avoir financé ce projet de recherche et avoir fourni les ressources nécessaires à la bonne réalisation de ce projet.

RÉSUMÉ

Ces dernières années, les solutions basées sur les technologies de l'industrie 4.0 deviennent de plus en plus accessibles aux PME. De meilleures stratégies de collecte de données sont développées et une augmentation des données disponibles est observée.

Pour certains secteurs industriels, les PME contribuent à une part majoritaire de la valeur économique produite. Les PME de ces secteurs industriels font face à des problématiques communes et évoluent dans des environnements similaires. L'étude des performances de productions des PME de ces secteurs industriels permet d'identifier une grande hétérogénéité de rendements et de qualités de production. Dans ces secteurs, il est possible de considérer les PME comme un réseau de PME décentralisé où les interactions entre chaque entreprise sont limitées.

Les systèmes d'information basés sur des technologies de l'industrie 4.0 sont présentés comme des outils potentiels permettant d'améliorer les performances de production des PME faisant partie de réseaux décentralisés. Aussi, l'étude, la comparaison et la valorisation des pratiques de gestion ainsi que des rendements de production de ces PME peuvent être réalisées par le développement d'outils collaboratifs au sein de réseaux de PME décentralisés.

Dans ce mémoire, nous proposons une méthodologie pour le développement d'une plateforme collaborative de partage de l'information pour visualiser et améliorer la performance des réseaux de PME. La méthodologie présentée couvre l'ensemble du processus de valorisation des données, de la définition du problème à la mise en production d'un outil collaboratif de visualisation et d'amélioration des performances des PME d'un réseau décentralisé.

La méthodologie proposée est constituée de 7 étapes décrivant la compréhension du contexte, la collecte des données, la préparation des données, la visualisation des données, l'amélioration des performances, le déploiement de l'outil collaboratif et la validation des critères de développement.

Finalement, la méthodologie proposée est appliquée sur un cas d'étude dans le secteur agroalimentaire de l'industrie fourragère canadienne. La plateforme collaborative développée permet de visualiser les performances de production, les techniques de gestion et les aspects spécifiques de chaque entreprise appartenant à un réseau de 240 exploitations réparties à travers le Canada. Ensuite, la plateforme permet de soutenir des pistes pour l'amélioration des performances de chaque entreprise à travers les meilleures pratiques qui sont identifiées dans des contextes similaires au sein du réseau. Le produit développé intègre 3 sources de

données différentes collectées à l'aide d'outils manuels et automatiques. Parmi les attributs collectés, les paramètres météorologiques tels que le taux de précipitation, la température, la pression ou bien les paramètres de gestions des récoltes tels que les rendements, les pratiques des agriculteurs sont introduites dans les bases de données du projet. Au total, plus de 70 paramètres sont collectés et intégrés à la plateforme de partage d'information développée dans le cadre de ce projet de recherche. L'ensemble de ces paramètres permet de produire 3 outils spécifiques décrivant les rendements de production, les bonnes et mauvaises pratiques ainsi que les caractéristiques des 240 exploitations du réseau d'entreprises étudié.

ABSTRACT

In recent years, solutions based on Industry 4.0. technologies are becoming more and more accessible to SMEs. As a result, better data collection strategies are being developed and an increase in available data is observed. For some industrial sectors, SMEs contribute a majority share of the economic value produced. SMEs in these industries face common issues and operate in similar environments.

The study of the production performance of SMEs in these industrial sectors allows us to identify a great heterogeneity of yields and production qualities. It is then possible to consider the SMEs of these sectors as a network of decentralized SMEs where the interactions between each company are limited.

Information systems based on Industry 4.0 technologies are presented as potential tools to improve the production performance of SMEs in decentralized networks. Also, the study, comparison and valorization of the management practices as well as the production yields of these SMEs can be achieved by the development of collaborative tools within decentralized SME networks.

In this thesis, we propose a methodology for the development of a collaborative information sharing platform to visualize and improve the performance of SME networks. The methodology presented covers the whole process of data enhancement, from the definition of the problem to the production of a collaborative tool for visualizing and improving the performance of SMEs in a decentralized network.

The proposed methodology consists of 7 steps describing the understanding of the context, the data collection, the data preparation, the data visualization, the performance improvement, the deployment of the collaborative tool and the validation of the development criteria.

Finally, the proposed methodology is applied through a case study in the agri-food sector of the Canadian forage industry. The collaborative platform developed allows to visualize the production performances, the management techniques and the specific aspects of each company. It then proposes ways to improve the performance of each company through the best practices identified in similar contexts within the network.

TABLE DES MATIÈRES

DÉDICACE	iii
REMERCIEMENTS	iv
RÉSUMÉ	v
ABSTRACT	vii
TABLE DES MATIÈRES	viii
LISTE DES TABLEAUX	xii
LISTE DES FIGURES	xiii
LISTE DES SIGLES ET ABRÉVIATIONS	xv
LISTE DES ANNEXES	xvi
CHAPITRE 1 INTRODUCTION	1
CHAPITRE 2 REVUE DE LITTÉRATURE	3
2.1 Problématique	3
2.2 Besoins et motivations	4
2.3 Le contexte spécifique des réseaux de PME	5
2.4 Techniques et indicateurs d'optimisation de la production	6
2.5 Architectures de données et solutions de mutualisation des données	7
2.5.1 Architectures de données	7
2.5.2 Mutualisation des données	8
2.6 Plateformes de partage de données existantes pour les PME	9
2.7 Synthèse	10
CHAPITRE 3 PROBLÉMATIQUE	11
3.1 Contexte industriel	11
3.2 Positionnement et problématique de recherche	12
CHAPITRE 4 MÉTHODOLOGIE DE DÉVELOPPEMENT D'UNE PLATEFORME DE PARTAGE DE DONNÉES POUR LES RÉSEAUX DE PME	13

4.1	Compréhension du contexte	14
4.1.1	Définition des termes du secteur	15
4.1.2	Analyse de la chaîne de valeur	15
4.1.3	Identification des besoins	16
4.2	Collecte des données	16
4.2.1	Identification des fonctionnalités du produit	17
4.2.2	Validation des besoins	17
4.2.3	Choix des données	18
4.2.4	Type et temporalité des données	18
4.2.5	Méthodes de collecte	18
4.2.6	Protocole de collecte	19
4.2.7	Architecture de collecte	19
4.3	Préparation des données	19
4.3.1	Structure de la base de données	20
4.3.2	Standardisation des colonnes	20
4.3.3	Nettoyage des données	21
4.3.4	Formatage des tables de données	21
4.3.5	Analyse et augmentation de la base de données	22
4.4	Outil de visualisation des données	22
4.4.1	Choix des indicateurs de performances	23
4.4.2	Analyse des besoins d'expériences utilisateurs	24
4.4.3	Conception de l'interface utilisateur	24
4.4.4	Test préliminaire de l'outil	24
4.4.5	Analyse des retours clients	25
4.4.6	Amélioration itérative	25
4.5	Optimisation des performances	25
4.5.1	Comparaison des performances des entreprises	26
4.5.2	Analyse des entreprises les plus et les moins performantes	27
4.5.3	Extraction des bonnes et des mauvaises pratiques	27
4.5.4	Présentation des indicateurs par l'outil	28
4.6	Déploiement	28
4.6.1	Anonymisation des données	29
4.6.2	Sécurisation des données	29
4.6.3	Gestion des accès	30
4.6.4	Maintenance et amélioration continue	30
4.7	Validation	30

4.7.1	Validation de l'utilisateur final	31
4.7.2	Validation des experts	31
4.7.3	Analyse des interactions	32
CHAPITRE 5	CAS D'ÉTUDE	33
5.1	Contexte et compréhension du marché: le secteur fourrager canadien	33
5.1.1	L'industrie fourragère	33
5.1.2	Déroulement d'un cycle d'exploitation	34
5.1.3	Rendement et qualité de production	35
5.1.4	Enjeux et besoins du secteur	35
5.1.5	Contexte industriel	36
5.2	Collecte des données	36
5.2.1	Protocole général de collecte	36
5.2.2	Collecte des données de production	38
5.2.3	Collecte des données de pratiques de gestion	40
5.2.4	Collecte des données de caractéristiques des sols	41
5.2.5	Collecte des données météorologiques	42
5.3	Préparation des données	45
5.3.1	Structure de la base de données	45
5.3.2	Nettoyage des noms de colonnes	47
5.3.3	Nettoyages des données	47
5.3.4	Formatage des données	51
5.4	Analyse et augmentation de la base de données	53
5.5	Outils de visualisation des données	54
5.5.1	Visualisation des pratiques de gestion	55
5.5.2	Visualisation des rendements de production	57
5.5.3	Visualisation de la santé des champs	58
5.5.4	Visualisation générale du réseau d'exploitants	60
5.6	Optimisation des performances	60
5.6.1	Outil de comparaison	61
5.6.2	Degrés jours	62
5.6.3	Perspectives d'améliorations	63
5.7	Déploiement	63
5.7.1	Anonymisation des données	63
5.7.2	Sécurisation des données	64
5.7.3	Gestion des accès	65

5.7.4 Amélioration continue	66
CHAPITRE 6 CONCLUSION ET RECOMMANDATIONS	68
6.1 Synthèse des travaux	68
6.2 Limitations de la solution proposée	69
6.3 Améliorations futures	69
RÉFÉRENCES	71
ANNEXES	75

LISTE DES TABLEAUX

Tableau 5.1	Paramètres collectés dans le cadre du projet	44
Tableau 5.2	Détection de valeurs aberrantes	48
Tableau 5.3	Traitement des valeurs manquantes	50
Tableau 5.4	Formatage des données en tables agrégées	52
Tableau 5.5	Paramètres de l'analyse de sol	59

LISTE DES FIGURES

Figure 4.1	Méthodologie générale	14
Figure 4.2	Processus régissant la compréhension du contexte	15
Figure 4.3	Organigramme du protocole de collecte des données	17
Figure 4.4	Organigramme du protocole de préparation des données	20
Figure 4.5	Structure d'une fonction produisant un nouvel attribut	22
Figure 4.6	Processus de visualisation des données	23
Figure 4.7	Processus d'optimisation des performances	26
Figure 4.8	Processus d'extraction des pratiques	27
Figure 4.9	Processus de déploiement de la plateforme	29
Figure 4.10	Processus de validation	31
Figure 5.1	Organigramme du protocole de collecte des données	37
Figure 5.2	Interface de collecte des données	38
Figure 5.3	Protocole de collecte de données de production	38
Figure 5.4	Outils de collecte des données	39
Figure 5.5	Interface de collecte des données de rendement	40
Figure 5.6	Protocole de collecte des données de pratique de gestion	40
Figure 5.7	Interface de collecte de pratique de gestion	41
Figure 5.8	Répartition des stations météorologiques du projet	42
Figure 5.9	Protocole de collecte des données météorologiques	43
Figure 5.10	Structure de la base de données	45
Figure 5.11	Architecture de la méthodologie	46
Figure 5.12	Visualisation des valeurs de rendements aberrants	48
Figure 5.13	Structure de la base de données formatées	53
Figure 5.14	Structure des variables générées à partir des données formatées	54
Figure 5.15	Représentation des pratiques de gestion	56
Figure 5.16	Représentation des rendements de production	57
Figure 5.17	Représentation de la santé des champs	58
Figure 5.18	Outils de présentation de la santé des sols	60
Figure 5.19	Courbe de Coupes en fonction des degrés jours d'une exploitation	62
Figure 5.20	Processus de protection des données	64
Figure 5.21	Architecture en valve choisie pour la visualisation des données	64
Figure 5.22	Présentation de différents niveaux d'accès	65
Figure 5.23	Gestion des versions de la plateforme	66

Figure A.1	Méthodologie complète de développement de l'outil collaboratif	75
------------	--	----

LISTE DES SIGLES ET ABRÉVIATIONS

ADF	Acid detergent fiber
API	Application Programming Interface
Ca	Calcium
CC+	Base de donnée de l'entreprise de conseils agronomiques
CPS	Cyber Physical System
CRISP-DM	Cross-Industry Standard Process for Data Mining
CSS	Cascading Style Sheets
DMME	Data Mining Methodology for Engineering applications
DPMS	Data Platform Management System
GDD	Growing Degree Day
GPS	Global Positioning System
HTML	HyperText Markup Language
IoT	Internet of things (Internet des objets)
N	Azote
NDF	Nutrient detergent fiber
OCDE	Organisation de coopération et de développement économiques
P	Potassium
PHP	Hypertext Preprocessor
PME	Petites et Moyennes Entreprises
R+	Logiciel agronomique
S	Soufre
T/H	Tonnes par hectare

LISTE DES ANNEXES

Annexe A	Méthodologie complète	75
----------	---------------------------------	----

CHAPITRE 1 INTRODUCTION

D’après l’OCDE (Organisation de Coopération et de Développement Économiques), la contribution des petites et moyennes entreprises (PME) au produit intérieur brut mondial se situe entre 50 et 60%. De plus, les PME représentent 99% des entreprises existantes et sont responsables de 70% des emplois [24].

Les PME sont des contributrices cruciales de l’économie mondiale et jouent un rôle très important quant au développement et à l’acquisition de nouvelles technologies. Au Canada, 41.9% du produit intérieur brut est réalisé par des PME [12]. Dans certains secteurs, les PME sont les contributrices principales, par exemple pour les domaines des services alimentaires (64 % de contribution au secteur), des soins et de la santé (86%), de la construction (65%), ou bien de l’agriculture (89%).

Ces 10 dernières années, la globalisation des systèmes d’information, la démocratisation des technologies d’industrialisation 4.0 ainsi que la numérisation des secteurs industriels ont permis d’accélérer le développement des technologies d’aide à la décision ainsi que des techniques d’apprentissage machine. Pourtant, il est démontré que les PME sont moins touchées par le développement de ces technologies et qu’elles sont moins connectées aux réseaux internationaux de connaissances [24]. Ainsi, au sein des secteurs composés majoritairement de PME, les performances de ces acteurs sont très variables et dépendent beaucoup de l’environnement de chaque PME.

Dans certains secteurs, la demande est supérieure à l’offre et la compétition est faible entre les acteurs du secteur. Les perspectives possibles pour répondre à la demande croissante sont alors l’amélioration des performances de production de l’ensemble des PME du secteur. Ces secteurs industriels peuvent former des réseaux décentralisés de PME où chaque entreprise fait face à des problématiques similaires dans des environnements pouvant différer. S’il existe certains outils permettant l’optimisation des performances de ces PME, le suivi, de la performance n’est pas individualisé et ne prends pas en compte les performances globales du réseau afin d’améliorer les performances de chaque entreprise du réseau.

L’émergence des technologies de l’industrie 4.0. dans ces environnements ainsi que le besoin croissant d’optimiser la qualité et les rendements de production de ces PME soulèvent alors la question de l’amélioration et de l’optimisation de ces réseaux décentralisés de PME qui contribuent pour une part importante à la richesse produite dans leurs secteurs.

Ces dernières années, le développement de systèmes d’informations permettant de gérer la

transmission de données décentralisées a permis l'apparition de solutions technologiques innovantes capables de monitorer et de faciliter l'échange d'informations et de ressources au sein de ces réseaux. Ces solutions technologiques ont trouvé leurs applications dans l'échange de produits financiers, la gestion de réseaux électriques ou encore la gestion de réseaux de distribution complexes. Elles permettent d'extraire des connaissances utiles pour chaque acteur du réseau à partir de l'étude globale des performances du réseau.

Les réseaux de PME étudiés dans le cadre de ce projet de recherche possèdent des caractéristiques similaires aux exemples présentés précédemment. Ainsi, l'utilisation de systèmes d'information et de plateformes d'échanges représentent des technologies prometteuses pour l'amélioration des performances de ces réseaux décentralisés de PME.

Dans cette optique, le présent mémoire propose une méthodologie pour le développement d'une plateforme collaborative visant l'optimisation des performances de production d'un réseau de PME.

Par la suite, le chapitre 2 présente une revue de littérature des problématiques associées aux réseaux décentralisés de PME et en particulier, les motivations et besoins d'un tel outil, le contexte spécifique des secteurs industriels étudiés ainsi que les solutions technologiques existantes pour les PME. Le chapitre 3 présente le contexte industriel dans lequel ce travail est effectué ainsi que la problématique traitée dans le cadre de ce projet de recherche. Le chapitre 4 propose une méthodologie pour le développement d'une plateforme collaborative visant à l'amélioration des performances des entreprises au sein d'un réseau de PME. Ensuite, le chapitre 5 applique la présente méthodologie à un cas d'étude spécifique: le secteur agroalimentaire et la production de fourrage canadien. Enfin, le chapitre 6 synthétise le travail effectué dans ce mémoire, apporte les limites de la méthodologie développée puis propose quelques perspectives d'évolution.

CHAPITRE 2 REVUE DE LITTÉRATURE

Le chapitre de revue de littérature présente un état de l’art des méthodes et des travaux de recherche utiles à la compréhension du contexte du présent projet de recherche. Il introduit dans un premier temps les principales problématiques liées aux réseaux décentralisés de PME puis présente les besoins et les motivations de tels réseaux. Ensuite, il décrit le contexte spécifique des réseaux de PME et présente succinctement les techniques existantes pour optimiser les performances d’une PME. Par la suite, ce chapitre présente les différentes solutions de mutualisation des données ainsi que les plateformes actuellement développées pour les réseaux de PME. Enfin une brève synthèse de la revue de littérature est effectuée afin d’identifier au mieux les besoins présents dans cette thématique de recherche.

2.1 Problématique

Dans un réseau décentralisé de PME, chaque entreprise agit selon ses propres besoins et vise à maximiser ses rendements économiques. L’objectif de telles entreprises est de minimiser les coûts associés à la production de biens et de services, du temps de travail ainsi que des ressources nécessaires au bon déroulement des activités opérationnelles [2].

L’analyse de certains secteurs industriels constitués majoritairement de PME montre que la communication entre ces entreprises est très limitée [2]. Pourtant, l’ensemble des PME contribuant à la production de biens ou de services dans ces secteurs peut être modélisé comme un réseau de PME décentralisé (grappe industrielle). Chaque entreprise fait alors face à des réalités ainsi que des problématiques similaires tout en possédant chacune des singularités économiques, géographiques ou environnementales différentes.

L’amélioration du partage d’information entre les entreprises d’un réseau représente une perspective possible pour l’optimisation des performances de production de chaque PME. Les technologies de sciences des données et de l’industrie 4.0 représentent des outils performants pour améliorer le partage de connaissances et de savoir-faire en fonction des besoins particuliers de chaque entreprise. De plus, le traitement et la mutualisation des données générées par ces entreprises représentent un autre aspect possible pour améliorer les performances de gestion de ces PME.

L’émergence des méthodes d’analyses de données telle que DMME [14] est étudiée dans ce mémoire afin de comprendre l’utilité des outils de partage de données appliqué aux réseaux de PME.

L'étude bibliographique effectuée dans le cadre de ce projet se concentre sur les 6 aspects suivants : tout d'abord, les besoins ainsi que les motivations justifiant la conception d'une plateforme de données pour les réseaux de PME sont présentés. Puis, les caractéristiques spécifiques des réseaux de PME sont étudiées dans le but de cerner le plus précisément possible leurs besoins. En 2.4, les techniques et les indicateurs d'optimisation de la production existants sont analysés afin d'identifier les outils manquant pour ces groupes d'acteurs économiques. En 2.5, les architectures de données existantes et les différentes solutions permettant de mutualiser des données pour un groupe de PME sont étudiées. En 2.6, les plateformes de données existantes visant à mettre en commun les données d'un réseau de PME sont analysées. Finalement, une synthèse est réalisée et permet d'identifier les perspectives d'amélioration développées dans ce mémoire.

2.2 Besoins et motivations

Cette sous-section a pour objectif d'identifier les besoins et les motivations des PME reliés à l'utilisation des technologies du numérique pour améliorer leurs performances de production.

Tout d'abord, les PME sont de plus en plus impliquées dans la transition numérique [38]. D'une part, de nombreux efforts sont actuellement développés afin d'accélérer l'implantation de systèmes de collecte de données au sein de PME [29]. D'autre part, de nouvelles solutions permettant le traitement de données pour des coûts abordables sont développées et mises à disposition [29]. Enfin, la diminution des coûts de production des capteurs, l'avènement des Systèmes cyber physiques (CPS), de l'internet des objets (IoT) et des technologies d'infonuagiques permettent aujourd'hui de bâtir des systèmes de collecte de données performants [18].

Les technologies d'industrialisation 4.0. sont aujourd'hui implémentées dans l'industrie de pointe et sont peu à peu transférées aux PME [30]. Cependant, l'utilisation des technologies d'industrialisation 4.0 pour la collecte de données au sein des PME produit un nombre considérable de données et contribue à la formation de réseaux décentralisés [21]. La comparaison des données de chaque entreprise du réseau n'est pas effectuée et aucune connaissance n'en est extraite.

Alors qu'au Canada, certains secteurs industriels comme l'agroalimentaire, la production manufacturière ou la construction sont majoritairement représentés par ces réseaux de PME. Le développement d'outils d'optimisation collaboratifs destinés à ces PME propose une perspective intéressante pour valoriser ces données [38]. La réalisation de tels outils expose de

nouvelles problématiques concernant le libre partage de données. En effet, si toutes ces entreprises souhaitent profiter des résultats de ces outils, leurs participations dans le développement sont très limitées, car elle nécessite la contribution de chaque entreprise en termes de donnée et peut révéler des informations stratégiques aux concurrents. Enfin, l'implantation de tels outils collaboratifs apporte une nouvelle perspective quant au partage des bonnes pratiques et au respect de l'anonymat des entreprises participantes. Ce mémoire tente donc de fournir un outil collaboratif impliquant les PME et permettant tant le développement individuel que global des acteurs de ce marché.

2.3 Le contexte spécifique des réseaux de PME

L'application des technologies de l'industrie 4.0 au contexte de PME est une problématique très étudiée dans la littérature [32]. Ainsi [8] et [21] ont développé des méthodes d'évaluation d'industrialisation 4.0 des PME. Ces méthodes ont pour objectifs d'évaluer les besoins de l'entreprise en termes d'industrialisation 4.0 et d'estimer la valeur ajoutée des technologies de l'industrie 4.0 à leur secteur d'activité. [4] fournissent un cadre d'implantation des technologies de l'industrialisation 4.0 et décrivent les principaux blocages pour l'installation de ces technologies au sein des PME.

Les faibles ressources financières, les rares utilisations des technologies de manufacture avancées et la spécialisation des produits développés sont les principaux facteurs qui rendent l'implantation des technologies d'industrialisation 4.0 difficile dans les PME. [13] propose donc une méthodologie d'application de l'industrie 4.0 en utilisant les systèmes cyber physiques (CPS) et les Big data afin de développer une plateforme de donnée générique. Cette plateforme est adaptée au contexte des PME et vise à fournir un cadre pour la collecte, la centralisation et la valorisation de données. Enfin, [27] développe des modèles théoriques basés sur l'infonuagique pour promouvoir les outils de cloud manufacturing et optimiser la communication entre chaque PME du réseau d'entreprise.

Les secteurs industriels constitués d'une majorité de PME possèdent des caractéristiques communes. Tout d'abord, dans ces réseaux d'entreprises, la répartition géographique des acteurs est très grande [36]. De plus, la communication entre chaque entreprise est limitée par leur grand nombre ainsi que par le caractère concurrentiel de leurs activités. Pourtant ces entreprises produisent une part majoritaire de la production totale. Chaque entreprise est propriétaire des données qu'elle récolte et la création d'initiative commune est difficilement réalisable. En effet, peu de standards adaptés au contexte des PME existent pour la collecte des données [30]. Ensuite, le manque de moyens et de connaissances limite l'utilisation des données quand elles sont récoltées [30]. La spécificité de chaque industrie représente une

barrière pour la mutualisation et la valorisation de ces données. Enfin, les données sont peu partagées, car elles décrivent la performance de l'entreprise et doivent respecter l'anonymat et la confidentialité de chaque acteur si elles sont partagées [8].

Le développement d'une plateforme capable de comparer les performances individuelles de chaque entreprise, de dégager les bonnes pratiques, d'établir un diagnostic de productivité tout en respectant le caractère confidentiel et anonyme des données de chaque entreprise s'intègre parfaitement dans ce contexte de PME. Si les PME possèdent de faibles capacités d'investissements dans la recherche et l'innovation, le retour sur investissement des projets doit être rapide et concluant. Contrairement à de grosses compagnies ou à des institutions publiques, les PME ne peuvent supporter le financement d'un projet de développement sur plusieurs années sans rentabilité à court terme. Cet argument influe ainsi sur la motivation des PME à participer à de tels projets de mutualisation de données.

2.4 Techniques et indicateurs d'optimisation de la production

Dans cette sous-section, les différentes techniques pour améliorer les performances de production sont identifiées à partir de la littérature. De plus, les grandes catégories d'indicateurs de performances sont étudiées.

Plusieurs outils visant l'amélioration des techniques de production ont été développés et présentés dans la littérature. [5] propose par exemple une méthodologie basée sur la production d'indicateurs de performance permettant de quantifier les performances de production d'une entreprise. La méthode développée par [5] utilise ainsi différentes catégories d'indicateurs répondant chacune à des besoins préétablis de l'entreprise. Ainsi, 4 catégories d'indicateurs sont distinguées: les indicateurs quantitatifs de production, les indicateurs de qualité, les indicateurs descriptifs relatifs à l'environnement et les indicateurs d'estimation de la sécurité. Ces 4 catégories permettent d'identifier les 4 aspects importants de l'optimisation des performances d'une entreprise.

Certaines autres méthodes prônent le développement empirique de ces indicateurs à partir du contexte industriel étudié [22]. Par opposition, [11] propose un cadre de développement holistique des systèmes de gestion de la performance. Lorsque plusieurs entreprises du même secteur industriel doivent être analysées, d'autres méthodologies complémentaires ont été développées. Ainsi, [35] propose un réseau d'indicateurs de performance ayant pour but d'établir les performances d'un réseau d'entreprise franchisées. La méthodologie proposée par [35] est divisée en trois phases d'optimisation. Dans un premier temps, des indicateurs sont construits afin de connaître la performance actuelle de l'entreprise. Ensuite, d'autres

indicateurs sont réalisés permettant l'amélioration des performances de production. Enfin, une catégorie d'indicateur est ajoutée pour mesurer l'impact des nouveaux indicateurs sur l'amélioration de la performance. Ce processus itératif permet alors d'apporter une visibilité sur l'ensemble des performances de production du réseau d'entreprise.

2.5 Architectures de données et solutions de mutualisation des données

Cette sous-section met en exergue les différentes architectures de données actuellement utilisées pour extraire et fournir de nouvelles connaissances à partir de données collectées. Les avantages et inconvénients des solutions de mutualisation de données y sont aussi étudiés.

2.5.1 Architectures de données

La conception d'une plateforme de données performante nécessite l'utilisation d'une architecture de données adaptée au contexte du problème étudié. Plusieurs solutions proposent des architectures de données permettant la collecte, la structuration, l'analyse et la présentation des données. Appliquées au contexte de PME, [23] propose 4 architectures adaptées à la conception des plateformes de données.

L'architecture dite de tiers multiple « Multiple tiers » [28] consiste à bâtir une plateforme à partir de plusieurs modules ayant chacun un rôle spécifique. Dans cette configuration modulaire, les étapes de présentation, de traitement, de stockage et de formatage des données sont réalisées par des modules tiers ne faisant pas partie du code source de l'application. Les avantages d'une telle architecture sont la rapidité de développement et l'indépendance des modules desservant les fonctionnalités de l'outil.

La deuxième architecture «broker pattern» [37] fait appel à un unique service de communication «broker» visant à gérer les communications de manière synchrone ou asynchrone entre les différents outils constituant la plateforme. Dans cette configuration, un outil commun gère les échanges entre les différents modules de communication. Cette architecture peut alors être interprétée à un réseau centralisé où une entité est garante de la communication des autres entités.

Ensuite, l'architecture en tableau ou «blackboard pattern» (blackboard pattern) [31] ou « design pattern» est constituée de trois modules. Le premier module a pour utilité d'enregistrer les informations globales de la structure de l'application. Puis, le second module est garant du maintien de la source d'information de l'application. Enfin, le dernier module permet de contrôler les échanges de données entre les deux premiers modules. Ce dernier exécute, configure et sélectionne les modules utiles à l'exécution de l'application.

Finalement, l'architecture par couche ou « *layered pattern* » [6] est constituée de plusieurs couches de traitement ayant chacune pour fonction d'effectuer une action sur le déroulement de l'application. L'ensemble des couches de traitement forme alors un processus permettant la réalisation de tâches plus complexes offerte par la plateforme de donnée. Les avantages d'une telle architecture sont la simplicité de développement et d'implantation. De plus, cette méthodologie permet de limiter le nombre de dépendances, car chaque couche est indépendamment codée. En addition, les étapes de test et de prototypages sont très faciles à réaliser ce qui permet de diminuer drastiquement le coût de développement de l'application par rapport aux autres architectures. Par opposition, cette solution est difficilement étendable au cours du temps, car le cadre de développement est limité. Enfin, le changement d'une couche peut avoir des impacts néfastes sur les couches subséquentes et le traitement de données en parallèle n'est pas possible. Le maintien d'une telle stratégie est donc difficile en opération, mais facilite grandement le prototypage.

De nombreuses autres architectures de plateforme peuvent être développées à partir de la combinaison de ces 4 solutions [19]. Ainsi, la combinaison de ces solutions permet de sélectionner certains avantages tout en limitant les inconvénients associés à certaines spécificités d'une architecture. Il existe d'autres contraintes liées au contexte spécifique des PME. Chaque entreprise possédant des standards différents pour maintenir leurs données, le risque d'erreur lors du processus de collecte des données est grand [9]. Ainsi l'architecture de données choisie doit être résiliente à ces erreurs. Ensuite, Les PME opèrent dans le même secteur industriel et ne doivent pas avoir accès aux données des autres partenaires. L'architecture de données choisie doit permettre de gérer la confidentialité des données ainsi que la sécurisation des différents niveaux d'accès.

2.5.2 Mutualisation des données

La mutualisation des données au sein d'un écosystème d'entreprise permet d'identifier et de segmenter des comportements d'intérêt et utiles à l'optimisation des performances de chaque entreprise participante [17]. En effet, le partage de données généralise et met en valeurs des problématiques avec une faible fréquence d'apparition au niveau de l'individu, mais repérable lorsque la population complète est étudiée [3].

À partir de ces observations, la mutualisation des données permet d'identifier les individus répondant au mieux à ces problématiques et d'apporter des solutions aux autres individus d'une population similaire. Si la mutualisation apporte une valeur quantifiable pour chaque individu, la gestion des communications, du partage et de la sécurité des données étudiées pose un défi important. En effet, en fonctions du contexte, plusieurs techniques ont été développées

dans la littérature pour gérer le partage de données au sein d'un écosystème d'individus. Les principaux défis associés à ce partage sont alors le respect de la confidentialité des données, le cryptage des données sensibles ainsi que la gestion et la limitation des accès.

[16] propose donc plusieurs solutions permettant de résoudre les défis de confidentialité des données. En effet, dans la visualisation des données, des filtres peuvent être installés afin de limiter l'accès aux sources de données aux différents niveaux d'utilisateurs. Ainsi, chaque utilisateur peut consulter uniquement les données qui le concernent. L'information peut alors être transférée du serveur à l'utilisateur, mais pas de l'utilisateur vers le serveur. De plus, l'hébergement du code source de l'application est seulement disponible sur le serveur d'hébergement et chaque utilisateur peut seulement visualiser une quantité restreinte de données. Ainsi, l'utilisateur n'a accès qu'à la dernière couche de l'architecture de la plateforme et la sécurité des données peut être garantie.

2.6 Plateformes de partage de données existantes pour les PME

Deux sous-catégories de plateformes de données peuvent être distinguées [25]. Tout d'abord, les plateformes de stockage et de gestion de données sont des applications permettant la réunion des données de plusieurs organisations. Elles peuvent inclure des couches de pré-traitement et de standardisation des données. Ensuite, les plateformes d'analyse de données visent, elles, à extraire des connaissances quantifiables du point de vue du marché apportant une réelle valeur ajoutée à partir des données collectées. La contribution de ces deux types de plateformes est nécessaire pour un processus d'extraction de connaissance à partir de données brutes.

[23] dresse un état de l'art des plateformes de partage de données dans le domaine de l'agriculture basé sur 45 études et spécifie le grand engouement actuel pour le développement de ce type d'outil. Les langages de programmation les plus utilisés pour bâtir ces plateformes sont JavaScript, PHP, HTML, CSS et Python. De plus, dans le cas de données géospatiales, [15] propose une approche de plateforme collaborative de donnée afin de mettre à disposition des données structurées pour de futurs outils. Les nombreux types de données disponibles dans le contexte de PME représentent ainsi un grand potentiel d'amélioration pour l'optimisation des performances de ces entreprises.

Si l'utilisation de plateformes collaboratives dans un contexte de PME apporte des avantages certains en minimisant les coûts associés au transport et à l'acquisition de matières premières, l'existence de méthodologie pour les développer ainsi que leur concrétisation reste très peu établie [13]. En effet, les blocages principaux à la concrétisation de ces ouvrages

sont des blocages opérationnels liés à la réalité des entreprises, des blocages environnementaux ainsi que comportementaux. Enfin, l'intégration de ces plateformes pose des défis d'interopérabilité, d'intégration, d'automatisation ainsi que de diffusion de l'information.

2.7 Synthèse

Le travail effectué dans le cadre de ce projet s'intéresse à l'amélioration des performances de production d'un réseau décentralisé de PME par le partage de connaissances. La réalité économique de ces grappes d'entreprises ainsi que l'environnement dans lequel elles évoluent soulèvent de nombreux défis de gestion des performances.

La problématique décrite dans ce chapitre est motivée par des besoins précis. En effet, l'évolution des techniques de collecte de données, la diminution du coût des méthodes de traitement de données ainsi que la nécessité de répondre à des enjeux similaires justifient le besoin de développer des outils de partage et d'analyses de gestion des performances pour les PME. De plus, le contexte dans lequel évoluent ces réseaux de PME fournit un bon cadre pour le développement de tels outils. En effet, dans ces secteurs, la demande est très importante et le défi de ces entreprises ne consiste pas à protéger leurs parts de marché. Le partage de données de production peut ainsi être envisagé sous certaines conditions précises.

Dans le cadre de cette revue de littérature, 4 catégories d'indicateurs de performances ont été identifiées pour représenter la gestion des performances d'une entreprise. Ces quatre catégories sont des indicateurs quantitatifs de production, des indicateurs de qualité, des indicateurs descriptifs et des indicateurs de sécurité. À la lumière de ces indicateurs, la section 2.5.1 identifie 4 architectures de données permettant le développement de plateformes de partage de données et relève les principaux avantages et inconvénients reliés à la mutualisation de données au sein d'un écosystème d'entreprise.

Enfin, la conception d'outils de partage et de mutualisation des données est déjà étudiée dans la littérature et la section 2.6 permet d'identifier les travaux déjà réalisés. Parmi ces plateformes existantes, certains outils se focalisent sur le stockage de données alors que d'autres plateformes tentent de produire des rapports analytiques sur des données déjà formatées.

Ainsi, il existe peu de solutions, d'après les connaissances rassemblées dans le cadre de ce travail, offrant une méthodologie de développement d'une plateforme collaborative couvrant de la collecte de données à la production de connaissances pour des groupes de PME. Ce présent mémoire tente ainsi de proposer une telle méthodologie visant à produire une plateforme de partage de données concentrée sur l'amélioration des performances de productions dans un contexte de réseaux de PME.

CHAPITRE 3 PROBLÉMATIQUE

Le présent chapitre a pour objectif d'établir la problématique de recherche. Pour ce faire, le contexte industriel dans lequel se situe ce projet est dans un premier temps décrit 3.1. Ensuite, à la lumière de la revue de littérature ainsi que du contexte industriel présenté, un positionnement est effectué sur la problématique choisie pour ce travail de recherche.

3.1 Contexte industriel

Le contexte industriel dans lequel s'inscrit ce travail de recherche est désigné par l'étude des réseaux décentralisés de PME. Plus spécifiquement, les réseaux étudiés sont des groupes constitués uniquement d'un grand nombre de PME. Un cas d'étude décrivant l'industrie fourragère canadienne est appliqué au chapitre 5. Les spécificités de ce secteur industriel décrivent bien le contexte industriel des réseaux de PME étudiés dans le cadre de ce travail de recherche. Ainsi, l'industrie fourragère vise à répondre aux besoins de production de fourrages destinés à l'alimentation animale. La majorité de la production de fourrages est réalisée par de petits exploitants faisant face à des contraintes de productions similaires et réparties sur un grand territoire.

Dans le secteur étudié, la demande est très supérieure à l'offre actuellement disponible. Aussi, chaque acteur du réseau n'est pas mis en concurrence directe avec les autres entreprises et cherche plutôt à optimiser ses performances de productions. Au même titre, la répartition géographique des entreprises du réseau est très grande au sein du territoire étudié. Dans ce contexte, la communication entre ces entreprises est très faible ou inexistante. De plus, les entreprises du réseau allouent peu de moyens pour développer des initiatives communes. Les acteurs du réseau travaillent dans un secteur industriel identique et font face à des défis de production, de gestion et d'optimisations similaires. Ainsi, les voies possibles d'amélioration des performances de ces entreprises sont l'optimisation des pratiques de gestion, des performances de production et des flux de communication. Enfin, dans ce contexte, la demande est croissante et les ressources permettant l'arrivée de nouveaux acteurs dans le réseau d'entreprise sont limitées.

Ainsi d'après les caractéristiques décrites le groupe de PME étudié peut être interprété comme un réseau décentralisé de PME.

3.2 Positionnement et problématique de recherche

La revue de littérature effectuée au chapitre 3 ainsi que le contexte industriel présenté dans ce chapitre permettent d'identifier plusieurs perspectives d'amélioration des performances d'un réseau de PME. Tout d'abord, si de nombreux efforts sont actuellement développés pour permettre à ces entreprises de numériser leurs opérations, peu de financement et de connaissances technologiques au sein de ces PME sont disponibles. Ensuite, le fort coût nécessaire à l'élaboration de solutions de mutualisation des données limite le développement de tels outils dans ce contexte spécifique d'un réseau de PME décentralisé.

L'état de l'art effectué dans la revue de littérature et plus spécifiquement les sections 2.4 et 2.5, 2.6 ont permis de démontrer que les solutions technologiques permettant le développement de tels outils sont fonctionnelles et actuellement utilisées dans d'autres domaines technologiques.

Le réseau de PME étudié est constitué d'un grand nombre d'entreprises réparties sur un vaste territoire. Les pratiques de gestion de ces entreprises sont similaires en termes des méthodologies d'application mais sont appliquées à des périodes différentes de l'année selon la position géographique de chaque entreprise. Les performances de production de chaque entreprise sont différentes et varient en fonction de la taille de chaque entreprise. Aucune entreprise présente dans le réseau de PME collecte actuellement les données nécessaires au développement d'une plateforme collaborative de données. Enfin, la majorité des entreprises du réseau ne dispose pas de fonds suffisants pour développer un outil d'aide à la décision permettant d'améliorer les performances de production d'une exploitation.

De ce fait l'objectif principal de ce mémoire est de développer une méthodologie permettant la conception d'une plateforme de partage d'information au sein d'un réseau décentralisé de PME.

À travers cet objectif général, une méthodologie de développement et d'implémentation d'un outil de visualisation des données pour les PME est proposée et un outil de comparaison pour les différentes entreprises de ce réseau est aussi développé. La faisabilité de cette méthodologie est finalement démontrée par sa mise en oeuvre sur un cas spécifique, appliqué à un secteur industriel: l'industrie fourragère canadienne.

CHAPITRE 4 MÉTHODOLOGIE DE DÉVELOPPEMENT D'UNE PLATEFORME DE PARTAGE DE DONNÉES POUR LES RÉSEAUX DE PME

Ce chapitre présente la méthodologie proposée pour le développement d'un outil d'aide à la décision pour les réseaux décentralisés de PME. Toutes les étapes nécessaires pour le développement d'une plateforme de partage y sont détaillées.

La méthodologie de développement, inspirée de la méthode DMME [14] et de la méthode CRISP-DM [7] est adaptée au contexte d'industrialisation 4.0 des PME. Celle-ci est composée de 6 étapes et débute par l'acquisition d'une expertise sur le domaine d'application des données (1). La collecte et la préparation des données sont ensuite réalisées en étapes (2 et 3). Puis, la présentation des performances individuelles est effectuée par le biais d'un premier outil de présentation (4). Une fois les données présentées, l'étape (5) optimise les performances de chaque entreprise. Enfin, l'outil collaboratif est déployé aux utilisateurs finaux (6).

Ce processus de développement est itératif. Ainsi, plusieurs boucles de rétroaction sont réalisées entre les étapes [2:3] et [1:4]. En effet, afin d'accélérer le processus de valorisation des données et de l'adapter aux besoins des PME, l'étape de présentation est utilisée pour identifier certaines valeurs aberrantes. De plus, l'optimisation des performances permet d'extraire des connaissances utiles à la compréhension du domaine. Ces connaissances sont ensuite réutilisées pour extraire de nouvelles données et fournir de nouveaux indicateurs.

Les étapes de la méthodologie utilisée sont présentées à la figure 4.1 et décrites dans cette section.

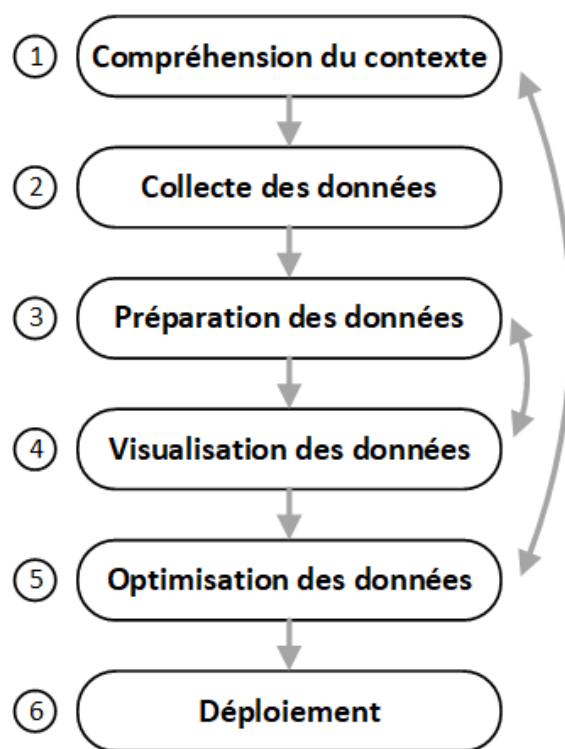


Figure 4.1 Méthodologie générale

Les mêmes étapes sont ensuite mises en oeuvre sur un problème industriel réel, en section 5. Cette mise en oeuvre a permis le développement d'une plateforme collaborative visant l'amélioration des performances de production d'un réseau d'entreprise.

4.1 Compréhension du contexte

Les objectifs de l'étape de compréhension du contexte sont de cerner les besoins du domaine ainsi que d'identifier les potentielles solutions technologiques répondant à ces besoins.

L'étape de compréhension du contexte est divisée en trois étapes successives. Tout d'abord, les termes et concepts principaux du secteur sont définis et étudiés. Ensuite, une analyse de la chaîne de valeur est présentée. Enfin, les principaux besoins du contexte industriel sont identifiés. À la suite de cette étape, le contexte englobant le secteur étudié est assimilé et les besoins associés à ce secteur sont établis. Ce processus est illustré à la figure 4.2.

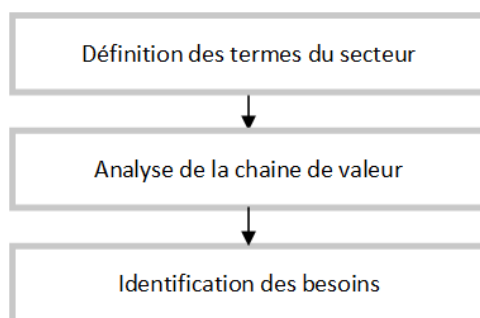


Figure 4.2 Processus régissant la compréhension du contexte

4.1.1 Définition des termes du secteur

Dans un premier temps, les termes principaux du secteur sont définis. En cela, le vocabulaire propre au domaine étudié est analysé et compris afin de pouvoir interagir aussi bien avec les experts du domaine qu'avec les opérateurs.

Lors de cette étape, il s'agit d'acquérir les connaissances de base du domaine. Ces connaissances peuvent être regroupées à partir de travaux de littérature et d'entrevues avec des experts du secteur. Les termes fréquemment utilisés par le secteur et permettant de juger de la qualité de production, de la productivité des entreprises ainsi que des principaux défis rencontrés par les acteurs du secteur sont définis.

4.1.2 Analyse de la chaîne de valeur

Ensuite, une analyse générale de la chaîne de valeur du secteur est effectuée. L'objectif de cette étape est de comprendre les principaux défis liés aux performances des entreprises du secteur. Cette étape permet d'identifier les éléments moteurs de cette industrie ainsi que les processus de création de valeur des entreprises du secteur étudié.

Lors de cette étape, le cycle de production est étudié afin d'identifier les facteurs limitants de ce cycle ainsi que de connaître les principales techniques théoriques et empiriques permettant l'amélioration des performances du secteur. De plus, à chaque maillon de la chaîne de valeurs, les intrants et les extrants sont étudiés. Il s'agit lors de cette étape de cartographier le processus de création de valeurs du réseau de PME afin d'en extraire les étapes clefs permettant d'identifier les besoins du secteur.

4.1.3 Identification des besoins

Enfin, à partir des deux étapes précédentes, les besoins du secteur sont identifiés et permettent d'introduire la phase de collecte de données. L'identification des besoins est réalisée à partir des spécificités du secteur, elle permet de dégager les caractéristiques propres du secteur comme la répartition géographique des entreprises dans le réseau, la vulnérabilité du secteur face à des variables économiques ou environnementales ou bien la facilité d'accès aux données pour les entreprises du réseau.

L'identification des besoins permet notamment de mettre en évidence une mauvaise distribution des ressources technologiques au sein du réseau d'entreprises. Cette étape sert alors à identifier ce qui pourrait être utile pour les acteurs du réseau et ce qui pourrait être implémenté pour améliorer les performances de gestion et de production de ces entreprises.

Les besoins du secteur peuvent être identifiés en échangeant avec les acteurs principaux du domaine, par l'utilisation de sondage ou d'entrevues et en consultant l'état de la littérature scientifique dans ce secteur. La finalisation de cette étape permet l'acquisition d'une connaissance suffisante du secteur pour mettre en place un protocole de collecte de données effectué à l'étape suivante 4.2.

4.2 Collecte des données

L'étape de collecte de données est effectuée au début de la méthodologie et permet d'établir les bases du projet. L'objectif est d'anticiper l'utilité du produit final à développer ainsi que de construire une architecture efficace de collecte de données. À partir des grandes lignes de ce produit, une stratégie de collecte de données peut être réalisée.

La méthodologie proposée dans le cadre de ce mémoire est présentée en figure 4.3.

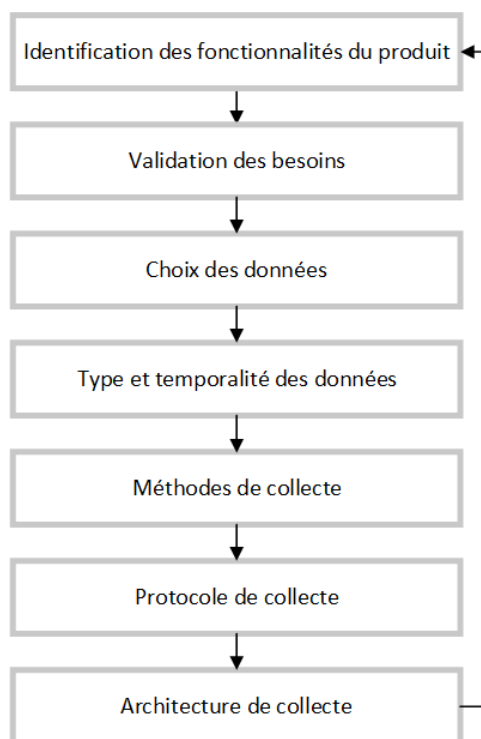


Figure 4.3 Organigramme du protocole de collecte des données

4.2.1 Identification des fonctionnalités du produit

Dans un premier temps, une identification des besoins et des fonctionnalités souhaitées dans le cadre du projet est réalisée en se basant sur la section 4.1 effectuée précédemment.

Au cours de cette étape, les utilisateurs finaux pour chaque outil à concevoir sont identifiés et leurs besoins énoncés. Des fonctionnalités de l'outil à concevoir sont définies pour chaque type d'utilisateurs identifié.

Ainsi, pour chaque besoin défini, plusieurs fonctionnalités peuvent être établies pour les différents types d'utilisateurs de l'outil. Ces fonctionnalités permettent ainsi de construire des maquettes de l'outil afin de les présenter et de les valider auprès d'un échantillon de futurs utilisateurs au cours de la prochaine étape.

4.2.2 Validation des besoins

Les besoins identifiés sont validés auprès d'un échantillon de futurs utilisateurs. Une première rencontre avec un échantillon de futurs utilisateurs est organisée afin de présenter les maquettes des outils qui seront développés. Cette première présentation permet de valider les besoins des utilisateurs et confirme la compréhension du cahier des charges du projet.

4.2.3 Choix des données

À partir des besoins validés, les données utiles à la réalisation de l'outil final sont déterminées. À partir des facteurs environnementaux impactant les performances des entreprises du réseau. L'ensemble des variables à collecter est alors rassemblé dans un document. Le type de données à récolter est déterminé par rapport à la compréhension du domaine ainsi qu'à l'aide des besoins du projet préétablis. À la fin de cette sous-étape, les données à collecter sont clairement établies.

4.2.4 Type et temporalité des données

Il s'agit maintenant, pour chaque attribut devant être collecté, d'identifier le type de données (temporelles, catégoriques, booléennes etc ...) ainsi que la vitesse à laquelle ces attributs seront collectés. Le type de données est défini au cours de cette étape comme le sujet des données et ne fait pas référence au format de données.

Chaque type de données a un impact sur les techniques de récolte devant être utilisées. Le choix du type de données de chaque attribut permettra ainsi de mieux établir le protocole général de collecte de données. De plus, la fréquence à laquelle chaque attribut est collecté doit être définie. En effet, les techniques développées pour collecter des données en temps réel sont plus coûteuses que des fréquences d'acquisition plus faibles. Le choix de la fréquence d'acquisition de chaque attribut influe donc sur le coût total du projet ainsi que sur le type d'information finale qui sera partagé par l'outil aux utilisateurs finaux.

4.2.5 Méthodes de collecte

Pour chaque attribut, une méthode de collecte doit être mise en place. Ainsi, cette étape détermine si des attributs doivent être collectés manuellement ou automatiquement et le type de capteur utilisé pour le processus de collecte est choisi.

Au cours de cette réflexion, les contraintes liées au contexte de collecte doivent être prises en compte. En effet, il est possible que le domaine étudié impose certaines saisonnalités dans la fluctuation des flux de données. Aussi, la grandeur du réseau sur lequel la collecte est réalisée peut impacter la qualité des méthodes de collecte. En effet, une grande répartition des acteurs sur le territoire, la faible disponibilité des réseaux de communication conventionnels ou le nombre d'individus dans le réseau peuvent entraîner des conséquences sur les choix de méthodes choisis.

4.2.6 Protocole de collecte

Une fois les méthodes de collecte de données choisies, des protocoles standardisés de collecte de données doivent être écrits permettant de décrire les processus de collecte de données sur la totalité de la durée du projet pour chaque attribut. Il s'agit donc de définir le protocole général de collecte de données du projet ainsi que d'assurer la traçabilité des attributs sur l'ensemble du cycle de vie des données collectées.

4.2.7 Architecture de collecte

Enfin, une architecture de données doit être conçue afin de déterminer l'emplacement, le stockage et le format des données collectées. Une fois l'architecture de données choisie, il est important de confirmer que l'ensemble des étapes valident bien les fonctionnalités et les besoins identifiés en première étape. À la fin du processus de collecte de données, une stratégie claire doit être établie prenant en compte toute la chaîne de collecte des données de l'identification des besoins et fonctionnalités de l'outil à l'architecture de collecte de données.

4.3 Préparation des données

En partant des données collectées, l'étape actuelle consiste à préparer les attributs collectés pour faciliter leurs futures utilisations dans des modèles de visualisation, de classification et de prédiction.

Pour ce faire, une structure de la base de données accueillant tous les attributs est choisie, le format de chaque nom d'attribut est standardisé et les données sont nettoyées. Ensuite, des tables agrégées de données sont construites afin de répondre aux problématiques établies dans la section de compréhension du contexte. Enfin, ces tables sont analysées et augmentées par de nouveaux attributs répondants aux besoins des clients.

Les attributs augmentés permettront aux clients de comprendre leurs pratiques de gestions, leurs performances actuelles de production ainsi que des attributs caractéristiques sur chaque entreprise. Ces étapes sont illustrées à la figure 4.4.

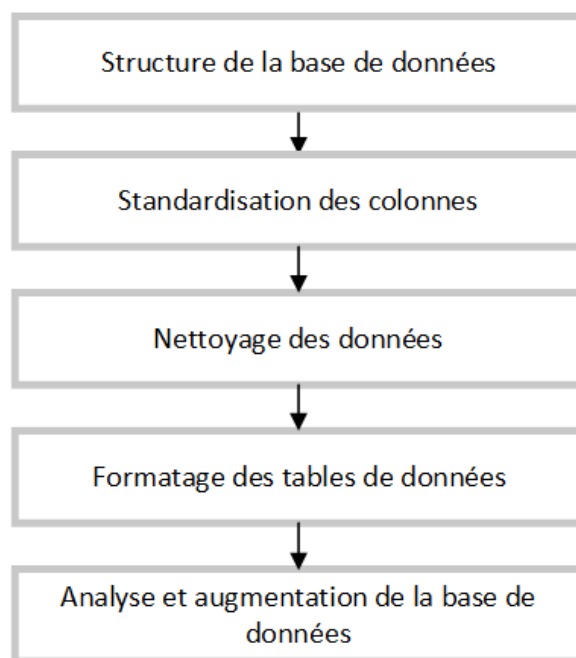


Figure 4.4 Organigramme du protocole de préparation des données

4.3.1 Structure de la base de données

Dans un premier temps, une structure de base de données est établie. Au cours de cette étape, trois requis doivent être validés. Tout d'abord, des tables de données comportant la liste des utilisateurs de la plateforme, leurs accès ainsi que la performance de chaque entreprise doivent être construits. Ensuite, les relations entre ces tables doivent être clairement identifiées. Enfin, la gestion des flux de communications entre les modules fonctionnels du code, les bases de données internes et des bases de données externes doivent être bâties.

En fonction des besoins du projet, les niveaux d'agrégations des données doivent être définis. Il existe plusieurs structures de base de données possible. Afin de choisir quelle structure est la plus appropriée, les avantages et inconvénients de chaque type de structure sont décrits par [31], [6] et [19]. Les singularités de chaque structure sont aussi décrites en section 2.5.1.

4.3.2 Standardisation des colonnes

Plusieurs standards peuvent être déterminés afin de définir un cadre d'utilisation de la base de données. Ces standards sont réunis dans un script permettant d'automatiser l'intégration de nouvelles données. La création de ces standards est réalisée lors de la création des processus de collecte de données et est réitérée à ce niveau de la méthodologie afin d'assurer la conformité de la structure de la base de donnée. Cette étape inclut ainsi la gestion des

données manquantes et des filtres de préparation des données. Plusieurs techniques pour gérer les données manquantes sont préconisées et peuvent être retrouvées dans ces recherches [33].

En fonction des besoins du projet, de la taille de la base de données ainsi que de la proportion des données manquantes, les valeurs manquantes peuvent être laissées telles quelles, remplacées par une valeur fixe, par une variable statistique représentante du projet ou encore par la moyenne des valeurs. En dernier recours, les données manquantes peuvent être corrigées en réalisant une nouvelle collecte de données lorsque c'est possible. Le choix de la méthode de standardisation des attributs dépend de l'importance de chaque attribut sur le contexte industriel ainsi que de la future utilisation de cet attribut dans l'outil à concevoir.

4.3.3 Nettoyage des données

Chaque attribut est nettoyé afin de garantir une bonne interprétation des résultats produits dans les étapes subséquentes. Le nettoyage des données aberrantes peut être effectué en se basant sur la méthode de préparation des données tirée de [26].

Les principaux types de données présents chez les PME sont des données de production temporelles et quantitatives. De plus, des données géospatiales ainsi que des données de pratique de gestions peuvent aussi être exploitées. Enfin, chaque entreprise possède des données statiques propres à ses caractéristiques. Les différentes techniques de nettoyage des données peuvent être basées sur des seuils fixes, des valeurs variables ou encore des valeurs statistiques. Le nettoyage des données est effectué par attributs et différentes stratégies de nettoyages peuvent être utilisées sur une même base de données.

4.3.4 Formatage des tables de données

L'information est formatée en plusieurs tables de données répondant chacune à un besoin analytique. Ces tables seront construites en fonction des besoins établis en section 4.2. Lors de cette étape, chaque attribut est renommé en fonction d'un standard établi dans le cadre du projet. Cette étape permet ainsi d'obtenir tous les attributs sous un même format pour un projet spécifique.

Par exemple, le nom de chaque attribut peut inclure la fréquence de collecte de l'attribut, les caractéristiques de l'attribut ainsi que l'information qu'il porte. Cette action assure une homogénéité de la base de données. De plus, elle permet d'identifier rapidement quels attributs sont modifiés dans le code source de l'outil conçu.

4.3.5 Analyse et augmentation de la base de données

À partir des attributs préparés, de nouveaux attributs sont générés en effectuant l'agrégation ainsi qu'en appliquant des fonctions sur des attributs collectés.

Pour chaque nouvel attribut ajouté, le code permettant de réaliser cette fonction doit respecter certains critères permettant l'indépendance de l'architecture de la plateforme de données. Chaque nouvelle fonction doit être basée sur des données brutes ou inclure ces propres données augmentées afin de garantir l'indépendance de développement de la plateforme ainsi que son aspect modulaire. Si le nouvel attribut fait appel à des bases de données externes, la fonction doit inclure l'appel des données externes. La figure 4.5 illustre la structure d'une bonne fonction.

```

Class Nouveaux indicateurs:
  Fonction 1
  def(Tables de données formatées):
    • Production des indicateurs de performances
    • Aggrégation des données par entités
  return (Nouvelle tables de données contenant les indicateurs)
  Fonction 2
  def(Tables de données formatées):
    • Production des indicateurs de performances
    • Aggrégation des données par entités
  return (Nouvelle tables de données contenant les indicateurs)

• Aggrégation des indicateurs dans un nouvelle table de données

```

Figure 4.5 Structure d'une fonction produisant un nouvel attribut

4.4 Outil de visualisation des données

L'étape de préparation des données étant effectuée, des outils permettant l'affichage et la mise en valeur d'indicateurs calculés sont construits.

L'objectif principal de cette section est alors d'apporter des connaissances supplémentaires aux entreprises participantes par la seule présentation des données récoltées. Cette étape permet dans un premier temps d'effectuer la validation des données auprès des experts du domaine. L'étape de présentation des performances actuelles permet de bâtir des indices de performance propres au domaine d'application et fournit un outil préliminaire d'aide à la décision pour les entreprises participantes.

Dans le contexte des PME, cette étape est divisée en trois tâches distinctes: l'analyse des

flux d'informations, l'étiquetage des pratiques de gestion et le suivi des performances de production. Ces trois étapes doivent ainsi permettre d'augmenter la vision stratégique des performances de l'entreprise ainsi que de décrire la qualité des pratiques de gestion.

Afin de réaliser ces objectifs, plusieurs étapes sont alors réalisées. Dans un premier temps, les indicateurs de performances permettant de décrire le contexte du réseau d'entreprise sont définis. En second temps, les besoins des entreprises sont repris à partir de l'étude du contexte 4.2 afin de bâtir une interface utile à l'environnement de ces PME. À la suite de la création des premiers tableaux, ces derniers sont testés sur un échantillon d'entreprise faisant partie du réseau. Enfin, les retours des utilisateurs tests sont analysés afin d'améliorer les tableaux générés. La figure 4.6 illustre ces étapes.

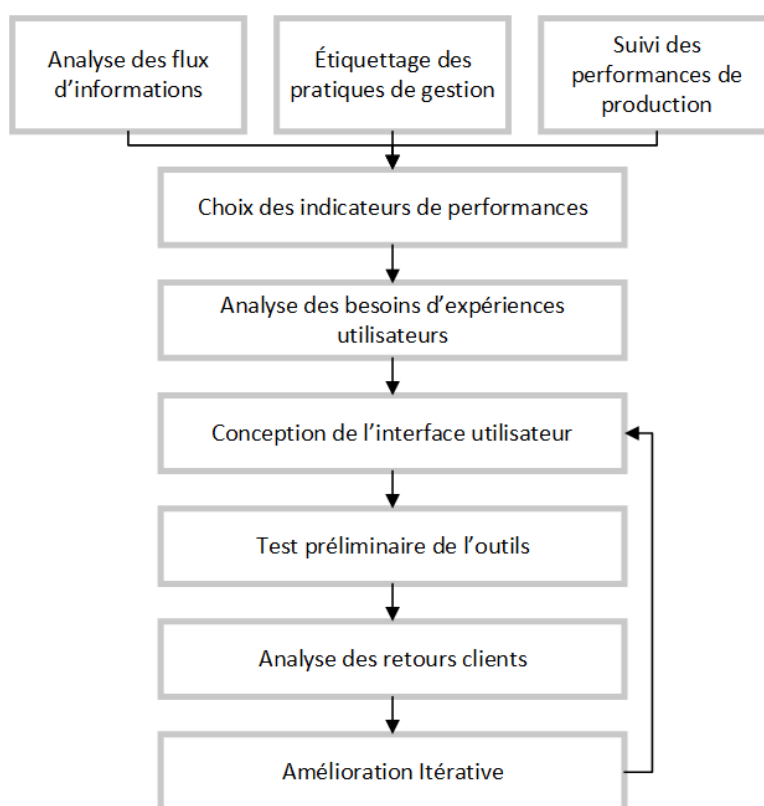


Figure 4.6 Processus de visualisation des données

4.4.1 Choix des indicateurs de performances

Afin de choisir les indicateurs de performances décrivant les besoins du réseau, une analyse des flux d'informations au sein d'une entreprise type du réseau ainsi qu'entre les différents acteurs du réseau est tout d'abord menée. En parallèle, les pratiques de gestion, de même que les performances de production, sont étudiées. À ce stade, il est important de connaître

les métriques caractéristiques propres au contexte industriel du réseau d'entreprise étudié. Par exemple, si le réseau d'entreprise est constitué d'entreprise de commerce en ligne, alors les indicateurs logistiques seront particulièrement intéressants. À partir de la cartographie des flux d'informations, des indicateurs des temps nécessaires à la réalisation de chaque processus peuvent par exemple être établis. Ensuite, l'étiquetage des pratiques de gestion peut permettre aux utilisateurs de l'outil de connaître les différentes étapes de chaque processus ainsi que leur qualité d'application. Enfin, l'étude du suivi des performances de production de chaque PME peut permettre de relier l'impact des flux d'informations et des pratiques de gestions sur ces performances.

4.4.2 Analyse des besoins d'expériences utilisateurs

La représentation de ces indicateurs sur une interface doit être étudiée afin de cibler clairement les besoins du réseau d'entreprise [10]. En addition, il est important de connaître l'environnement dans lequel seront consultés les outils, le type d'utilisateur qui consultera ces outils, la fréquence d'actualisation nécessaire de ces tableaux ainsi que le nombre de corps de métier différents qui utiliseront ces outils. Ensuite, l'information doit être assez synthétisée et facile d'accès. À partir de l'analyse des besoins de l'expérience client, des interfaces utilisateurs peuvent être conçues permettant d'afficher l'information utile à chaque type d'utilisateur sous un format répondant aux contraintes de consultations. Ces contraintes sont généralement liées à l'environnement dans lequel évolue l'utilisateur final.

4.4.3 Conception de l'interface utilisateur

Pour la conception de l'interface utilisateur, il est préconisé d'utiliser des plateformes de visualisation de données déjà développées par des sous-traitants. Ces solutions ont l'avantage d'être robustes et évolueront avec le temps. Dans le cadre du cas d'étude présenté au chapitre 5 un exemple basé sur l'outil de visualisation de données Tableau est présenté et appliqué au problème industriel.

4.4.4 Test préliminaire de l'outil

Les tableaux produits sont testés sur un échantillon d'utilisateurs. Il faut s'assurer lors des tests que l'information affichée répond aux besoins de l'utilisateur et que ces outils soient compris et simples d'utilisation pour l'utilisateur final. Lors de cette étape, un outil de rétroaction est mis en place afin de récupérer les retours des utilisateurs et d'améliorer l'outil pour les futures itérations de développement.

4.4.5 Analyse des retours clients

À partir des retours des utilisateurs tests, les améliorations possibles doivent être déterminées, priorisées puis implémentées. La sélection des améliorations à développer peut être effectuée en fonction des ressources nécessaires, du nombre de retours identiques sur une demande spécifique, de l'apport potentiel de cette amélioration pour les utilisateurs ainsi que du coût de cette amélioration. Une fois sélectionnées, les améliorations sont priorisées en fonction de leur facilité de développement et de leurs valeurs potentielles pour les utilisateurs. Ainsi, les premières améliorations à effectuer sont des améliorations pouvant être développées rapidement pour un coût faible et qui apporte une grande valeur aux utilisateurs.

4.4.6 Amélioration itérative

À la fin de cette étape, des tableaux sont construits permettant d'afficher des informations utiles aux utilisateurs du réseau d'entreprise. Leur utilité est confirmée par un échantillon de la population. La prochaine étape consiste alors à optimiser les performances de ces entreprises en utilisant des outils d'intelligences artificielles ainsi qu'en produisant de nouveaux indicateurs de performances.

4.5 Optimisation des performances

À la suite de la présentation des performances de base des entreprises, des modules de quantification, d'estimation et de comparaison sont ajoutés lors de cette étape afin d'optimiser les performances de chaque entreprise.

Des outils de partage des bonnes pratiques de gestion basées sur des modèles théoriques et empiriques existants sont tout d'abord implémentés. Puis, d'autres modules de comparaison des bonnes pratiques présentes dans le réseau d'entreprise sont développés afin de classer les performances de chaque entreprise dans des groupes par rapport à ses caractéristiques spécifiques.

À la fin de cette sous-étape, des groupes d'entreprises de performances et de caractéristiques similaires sont obtenus. Ces similarités dépendent du domaine étudié et peuvent être basées sur la taille, la situation géographique et d'autres attributs présents dans la base de données. Les pratiques de gestion de ces groupes peuvent être analysées et implémentées par l'entremise d'indicateurs de performance dans la plateforme. Ces indicateurs de performance sont finalement partagés pour chaque entreprise possédant des caractéristiques similaires. Ces étapes sont décrites par la figure. 4.7

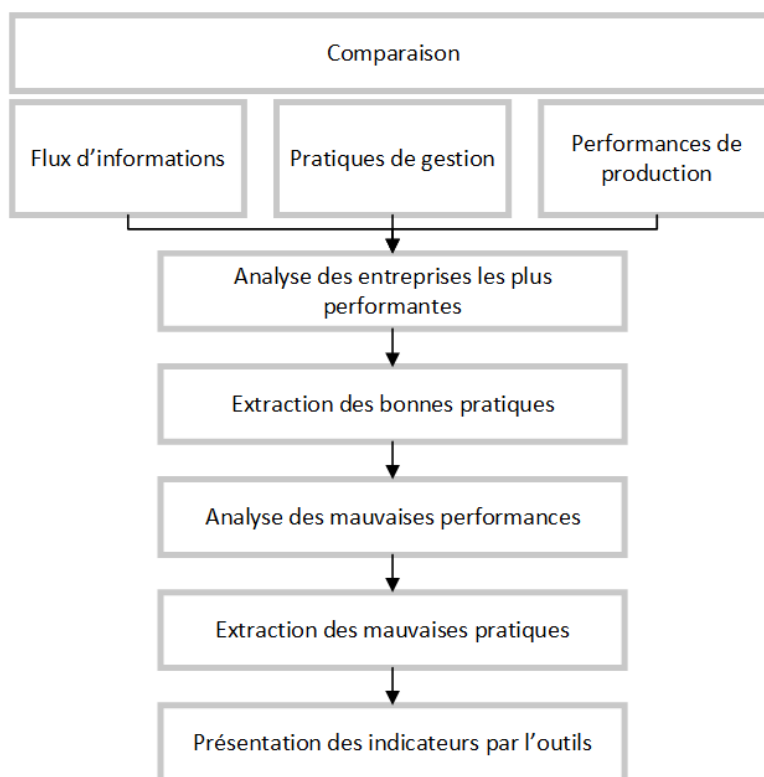


Figure 4.7 Processus d'optimisation des performances

4.5.1 Comparaison des performances des entreprises

Dans un premier temps, la comparaison des performances des différents acteurs du réseau est réalisée. L'étape 4.4 a permis de présenter les performances actuelles de chaque entreprise du réseau de PME. À partir de ces performances, l'objectif de cette étape est de comparer ces performances et de fournir un nouvel indicateur à chaque entreprise afin d'évaluer sa performance par rapport à d'autres acteurs du réseau.

En fonction du secteur étudié, cette comparaison peut être effectuée en fonction de critères géographiques, de critères de vraisemblances, de tailles d'entreprises ou bien de critères économiques. Ainsi, si cette comparaison est effectuée selon un critère géographique, l'indicateur développé peut permettre à chaque entreprise d'évaluer ses performances de productions par rapport aux 5 entreprises les plus proches. Afin d'anonymiser les informations et selon le secteur étudié, l'indicateur conçu peut afficher la moyenne de ces 5 entreprises ainsi que le classement de l'entreprise considéré dans ce groupe d'entreprise.

4.5.2 Analyse des entreprises les plus et les moins performantes

La comparaison des performances des entreprises ayant été réalisée dans la section précédente, une analyse des entreprises les plus performantes est ensuite effectuée. Cette étape permet d'extraire les facteurs ayant le plus d'influence sur la performance des entreprises. Il s'agit donc dans cette étape d'identifier et de regrouper les entreprises les plus performantes ainsi que les entreprises les moins performantes. Ensuite, l'analyse des pratiques de gestion ainsi que des attributs caractéristiques de chaque entreprise comprise dans ces deux groupes est effectuée.

4.5.3 Extraction des bonnes et des mauvaises pratiques

À partir de l'analyse de ces acteurs, une extraction des bonnes pratiques ainsi que des mauvaises pratiques peut alors être effectuée. Pour ce faire, des techniques d'apprentissage machines peuvent être utilisées telle que des méthodes Apriori. Une liste non exhaustive des méthodes pouvant être utilisées pour effectuer ces analyses est décrite par [1].

L'objectif de cette étape est de pouvoir dégager des recommandations permettant d'améliorer les performances des entreprises du réseau en fonction des pratiques extraites à l'étape précédente. Chaque PME au sein du réseau peut alors bénéficier d'un nouvel indicateur recommandant de bonnes pratiques basé sur les performances des entreprises les plus performantes et en fonction de critères de similarité.

L'extraction de ces pratiques est illustrée en figure 4.8. cette figure présente ainsi, le format des vecteurs utilisés pour l'extraction des bonnes et des mauvaises pratiques.

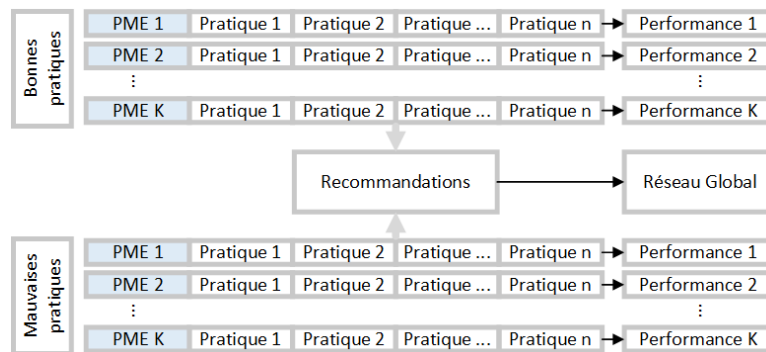


Figure 4.8 Processus d'extraction des pratiques

4.5.4 Présentation des indicateurs par l'outil

Enfin, ces analyses produisent de nouveaux indicateurs permettant de communiquer de nouvelles informations aux partenaires industriels du réseau de PME. En addition, des indicateurs basés sur des techniques empiriques et théoriques peuvent être utilisés pour ajouter de nouvelles analyses à la plateforme de données. À la fin de cette étape, les acteurs du réseau disposent d'un outil leur offrant des recommandations utiles à l'amélioration de leurs performances de productions. La prochaine étape de cette méthodologie consiste alors à réaliser une stratégie de déploiement de l'outil développé.

4.6 Déploiement

L'étape de déploiement est effectuée lors de la présentation des données ainsi qu'à la suite de l'optimisation des performances unitaires de chaque entreprise. Au cours de cette étape, une attention particulière est portée sur la préservation de l'anonymat des partenaires. Aussi, la confidentialité des données utilisées doit être respectée. Enfin, les indicateurs de performances développés pendant l'étape d'optimisation des performances doivent garantir la sécurité et l'intégrité des données au cours du déploiement.

Cette étape se décompose en 3 axes. Dans un premier temps, une attention particulière doit être attribuée à l'anonymisation des données ainsi qu'à la confidentialité des caractéristiques de chaque partenaire. Ensuite, un plan développant la gestion et la sécurité des accès pour chaque partenaire est développé. Enfin, un plan de maintenance est réalisé afin de prendre en charge les futures améliorations possibles dans l'architecture de la plateforme. Ce processus est illustré à la figure. 4.9

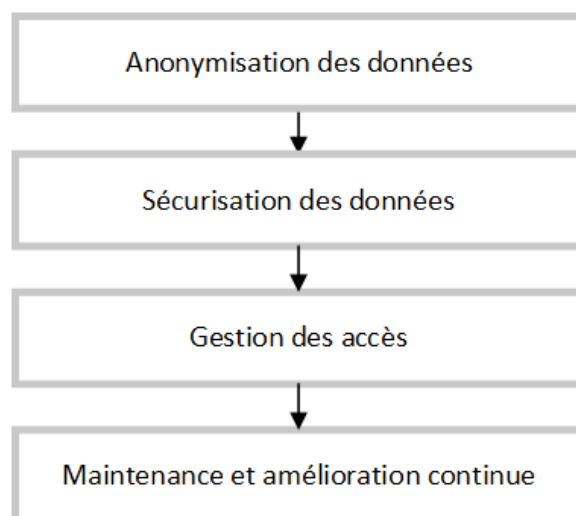


Figure 4.9 Processus de déploiement de la plateforme

4.6.1 Anonymisation des données

Dans un premier temps, des techniques d'anonymisation des données peuvent être utilisées pour limiter l'identification des partenaires. [20] effectue une revue de littérature sur les techniques d'anonymisation et de sécurisation des données. Les noms des colonnes peuvent être cryptés et les informations caractéristiques des partenaires remplacées par des identifiants.

En fonction du secteur industriel, une anonymisation relationnelle ou structurelle peut être effectuée sur les bases de données utilisées. D'une part les données sont anonymisées afin de supprimer toute information de relation entre des attributs et des entreprises. D'autre part, la structure de certains attributs peut être altérée afin d'éliminer l'identification des performances spécifiques des PME [20] décrit les 5 principales techniques d'anonymisation qui sont la généralisation, la suppression, la permutation, la perturbation et l'anatomisation.

En fonction des besoins spécifiques du secteur étudié et de chaque attribut, une combinaison de ces techniques peut être utilisée pour anonymiser certains indicateurs de performances de la base de données.

4.6.2 Sécurisation des données

Dans cette étape, une architecture permettant la sécurisation des données en fonction du contexte industriel du réseau d'entreprise est conçue. Plusieurs solutions sont possibles telles que les architectures en vannes permettant seulement aux partenaires de consulter les données qui le concernent sans pouvoir accéder aux données des autres partenaires. En fonction du

type d'architecture choisie, les données peuvent être cryptées à la source ou bien lors du transfert entre la base de données et l'utilisateur. Finalement, des outils de visualisation peuvent être utilisés afin de limiter les droits d'accès des utilisateurs à la base de données.

4.6.3 Gestion des accès

En troisième étape, un système de gestion des accès doit être pensé en fonction des besoins des utilisateurs. Concrètement, une instance doit être utilisée pour stocker les données de connexions des utilisateurs. Cette instance doit être en mesure de relier des identifiants de connexions à chaque partenaire ainsi que des niveaux d'accès au sein d'un même partenaire. Ces solutions peuvent être basées sur des systèmes déjà développés.

La mise en place d'un système de gestion des accès permet de limiter la visibilité de certains utilisateurs à leur domaine d'application. L'objectif d'un tel système est aussi de suivre le nombre d'utilisateurs pouvant utiliser l'outil développé et de gérer leurs utilisations de la plateforme. En général, ce système est constitué d'une table de données contenant les identifiants de connexion, les historiques de navigation sur l'outil ainsi que des liens permettant d'identifier les données accessibles à chaque utilisateur.

4.6.4 Maintenance et amélioration continue

Enfin, dans l'architecture et la structure du code source de l'application, une stratégie de maintenance ainsi que d'introduction de nouveaux modules actifs doit être conçue en fonction de l'architecture de la plateforme choisie 2.5.1. Ainsi, si l'architecture choisie est basée sur l'association de différentes couches fonctionnelles telle qu'effectuée dans le cas d'étude 5, alors la structure de déploiement doit comprendre l'implémentation de nouvelles couches. À la suite de l'étape de déploiement, les entreprises disposent d'un outil collaboratif capable d'optimiser leurs performances de production. Finalement, d'autres perspectives peuvent être développées en se basant sur cette méthodologie. Ainsi, les prochaines étapes visent à l'amélioration des performances globales du réseau d'entreprise en utilisant des outils d'apprentissage machine. La prochaine section présente alors la validation de la méthodologie utilisée pour concevoir un outil collaboratif d'amélioration des performances d'un réseau d'entreprise.

4.7 Validation

L'objectif de l'étape de validation permet de vérifier si les besoins identifiés à l'étape de compréhension du contexte sont bien comblés par les fonctionnalités de l'outil.

Afin de valider ces requis, 3 sous-étapes sont réalisées. En premier lieu, un processus est mis en place afin de valider les requis de l'utilisateur final. En l'occurrence, ces utilisateurs sont les employés et les dirigeants du réseau de PME. Ensuite, un second processus de vérification est réalisé afin de valider l'utilité de l'outil auprès des experts du domaine. Ces derniers valident à nouveau la pertinence des indicateurs ainsi que leur implémentation dans l'outil. Enfin, l'utilisation de l'outil est sondée auprès de tous les utilisateurs afin d'analyser leurs interactions avec la plateforme.

Ces différentes étapes sont illustrées à la figure 4.10.

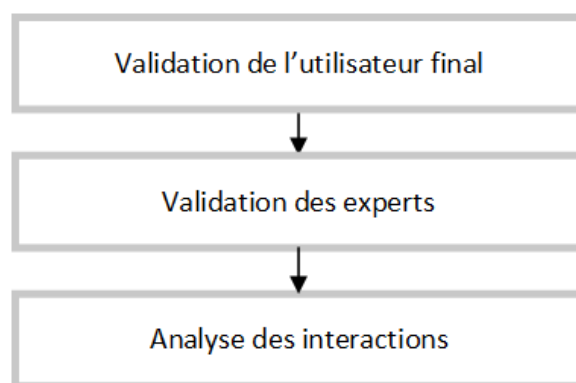


Figure 4.10 Processus de validation

4.7.1 Validation de l'utilisateur final

Une validation de l'outil par l'utilisateur final est dans un premier temps effectuée. Ensuite, une présentation est organisée afin de faciliter la prise en main de l'outil. Lors de cette présentation, chaque outil développé est décrit. De plus, les requis identifiés initialement y sont repris afin de justifier les choix de développement. Au cours de cette présentation, l'ensemble des étapes du projet y sont présentées. L'identification des besoins du secteur, le cahier des charges ainsi que les étapes de développement du projet sont rappelés. Enfin, le protocole d'utilisation de l'outil est présenté et une séance de question est organisée afin de s'assurer de l'adoption de l'outil par les utilisateurs finaux.

4.7.2 Validation des experts

Ensuite, les indicateurs présentés sont validés auprès d'un comité d'experts. Cette action permet de confirmer une dernière fois l'utilité de l'outil pour répondre aux besoins du réseau de PME ainsi que de valider la rigueur des indicateurs développés. L'outil est présenté au comité d'expert afin de valider sa cohérence ainsi que de promouvoir son utilisation au sein

du réseau de PME.

4.7.3 Analyse des interactions

Enfin, les perspectives possibles d'amélioration de l'outil sont sondées auprès des utilisateurs. Lors de cette étape, les différentes interactions entre la plateforme et les utilisateurs telles que le nombre de consultations par utilisateur et les tableaux les plus consultés sont analysées. Ces trois sous-étapes permettent de valider l'utilité de la plateforme auprès des utilisateurs. Ces actions permettent d'achever la méthodologie présentée dans le cadre de ce projet de recherche.

Le déroulement complet de cette méthodologie aboutit à la conception d'une plateforme collaborative destinée à l'amélioration des performances des PME au sein d'un réseau décentralisé. Le chapitre 5 se base ainsi sur cette méthodologie pour développer une plateforme collaborative afin d'améliorer les performances des PME d'un réseau décentralisé de l'industrie fourragère canadienne.

CHAPITRE 5 CAS D'ÉTUDE

Ce chapitre montre la mise en oeuvre de la méthodologie présentée dans le chapitre 4. Chaque étape de la méthodologie est reprise et concrétisée à travers un exemple réel: l'industrie fourragère canadienne.

Au cours de ce chapitre, une étude du domaine fourrager est dans un premier temps effectuée. Ensuite, le processus de capture de données est décrit. Puis, les outils nécessaires à la préparation de la base de données sont présentés. En sous-section 5.4, la base de données est augmentée par la création d'indicateurs de performance. Enfin, l'outil produit dans le cadre de ce projet est présenté ainsi que son implémentation et son déploiement. Une synthèse des étapes réalisées et des impacts du projet est alors présentée.

À l'exception de l'application de collecte de données pour Ipad réalisée par l'entreprise partenaire et illustrée en figure 5.5, l'ensemble des travaux présentés au cours de ce chapitre est réalisés par l'étudiant.

5.1 Contexte et compréhension du marché: le secteur fourrager canadien

Cette section a pour objectif de décrire le contexte associé au secteur industriel de l'industrie fourragère canadienne et plus spécifiquement de l'exploitation de luzerne au Canada. Elle définit le secteur de l'industrie fourragère canadienne puis explicite le déroulement d'un cycle d'exploitation d'un champ de luzerne. Elle décrit les défis liés à l'obtention de bons rendements et d'une bonne qualité de culture afin de présenter les enjeux et besoins de ce marché. Enfin, le contexte industriel spécifique à ce projet de recherche y est présenté.

5.1.1 L'industrie fourragère

L'industrie fourragère est un secteur industriel qui a pour objectif de produire des fourrages destinés à la consommation animale. Les fourrages sont utilisés pour nourrir des bovins, des équidés ou des ovins afin de produire du lait et de la viande.

Au Canada, l'industrie fourragère est majoritairement représentée par la production de luzerne. La luzerne est une légumineuse exploitée pour sa haute teneur en protéine. 89.5% de la production totale de luzerne est réalisée par de petits producteurs [12].

Les exploitations sont réparties sur la totalité du territoire canadien et l'on en retrouve une concentration plus importante au Manitoba, au Saskatchewan et au Québec. L'ensemble de

entreprises productrices de luzerne forme un réseau décentralisé où chaque ferme fait face à des paramètres environnementaux très différents. Souvent, l'utilisateur final du fourrage est aussi le producteur.

Les producteurs sont assistés au cours des saisons par un réseau d'experts agronomiques ayant pour objectif d'optimiser les performances de productions des champs exploités. Actuellement, les méthodes d'optimisations utilisées sont basées sur des connaissances empiriques et des modèles mécanistes [34]. Néanmoins, l'utilisation de modèles théoriques n'est pas démocratisée sur l'ensemble du réseau d'entreprises exploitantes. De plus, il existe peu d'outils permettant de faciliter la consultation de modèles théoriques par le producteur.

Le manque de standard pour la collecte de données agricole à l'échelle nationale complique le partage de données entre producteurs et le bon déroulement d'initiative commune visant à mutualiser des données agricoles. L'objectif de ce cas d'étude est d'appliquer la méthodologie présentée au chapitre 4 afin de bâtir un outil collaboratif d'aide à la décision.

L'outil développé à l'aide de la méthodologie proposée vise à mutualiser des données récoltées au sein d'un réseau de 240 PME du secteur fourrager. L'objectif final est d'améliorer la visibilité des producteurs sur leurs exploitations, d'améliorer leurs performances de gestions ainsi que de maximiser la productivité des exploitations en termes de rendement et de qualité de production.

5.1.2 Déroulement d'un cycle d'exploitation

Le cycle d'exploitation d'un champ de luzerne est réalisé sur une durée de 3 à 5 saisons. À la suite de ce cycle, il est impossible de semer à nouveau de la luzerne sur ce champ puisque cette légumineuse possède des propriétés d'auto toxicité. La luzerne produit des protéines tuant les jeunes pousses de luzernes. Ces protéines se dégradent sur une période de 2 ans et il est donc impossible de cultiver à nouveau de la luzerne dans les 2 années suivant la fin d'exploitation d'un champ de luzerne.

Au cours du cycle d'exploitation, la luzerne est récoltée de 3 à 5 fois par saison. Ces coupes sont réalisées au Canada de mai à octobre. Chaque coupe produit une récolte de différents rendements et qualités. Au cours des années d'exploitation, les champs perdent en persistance du fait de l'épuisement des ressources et du taux d'auto toxine dans les sols. Ainsi, les récoltes des années 4 et 5 sont généralement de rendements plus faibles que celles des premières années.

Afin de maximiser les rendements et la qualité des fourrages produits, les exploitants disposent aujourd'hui de plusieurs outils pouvant avoir un impact sur le résultat final. Un bon entretien des sols, un bon calendrier de récolte, un traitement optimal des nuisibles ainsi que

des pratiques de gestion optimales peuvent avoir un impact conséquent sur les rendements et la qualité de la luzerne produite au cours du cycle d'exploitation.

5.1.3 Rendement et qualité de production

La luzerne produite étant destinée à la consommation animale, le suivi de la qualité des récoltes est un aspect important pour l'industrie fourragère.

Avant chaque récolte, les agronomes collectent un échantillon de luzerne et effectuent une analyse de fourrage. Ces analyses permettent de décrire explicitement la qualité de la luzerne récoltée. Une luzerne de bonne qualité, consommée en quantité suffisante par l'animal, transmet un grand apport en protéine. À l'inverse, une luzerne de mauvaise qualité est consommée en excès par l'animal et est mal assimilée.

L'importance d'obtenir de bons rendements pour le producteur est de ne pas manquer de fourrage pour ses animaux lors de l'hiver. Un manque de fourrage entraîne des dépenses supplémentaires ainsi qu'une diminution des profits économiques de l'exploitant. À l'inverse, une surproduction n'est pas forcément néfaste pour le producteur puisque l'excédent peut être revendu à d'autres producteurs.

5.1.4 Enjeux et besoins du secteur

Les caractéristiques de l'industrie fourragère sont les suivantes. Tout d'abord, l'ensemble de la production de luzerne canadienne est réalisé par de petits producteurs.

Les producteurs ont l'organisation typique d'une PME. Chaque exploitation est gérée par un petit nombre de personnes avec des moyens financiers limités. L'ensemble de ces fermes, à l'échelle du Canada, est très dispersé géographiquement et il existe peu de communications entre les différents acteurs. Une grande partie de ces PME font face à des problématiques communes et mettent en place différentes stratégies de résolution ayant une efficacité variée. Il existe un réel besoin pour ces PME d'augmenter les flux de communication.

Le suivi des performances de production est sujet à des problèmes de surproduction pour certains producteurs et de sous-production pour certains producteurs. L'augmentation des canaux de communication peut donc être considérée comme une bonne solution pour réaliser la distribution des surplus de luzerne.

5.1.5 Contexte industriel

Pour les besoins de ce projet, 240 fermes sont sondées. Afin de récolter un maximum d'information sur les performances de ces exploitations, de multiples capteurs sont installés et collecteront des données pour les besoins du projet pendant 5 ans (2021-2025). Le travail décrit dans cet article est basé sur les données récoltées au cours la première année (2021). La collecte des données est réalisée par l'entreprise Jasons système fourrager, le stockage des données est pris en charge par l'entreprise LOGIAG et la préparation des données et des étapes de valorisation sont effectuées par les auteurs de ce projet.

5.2 Collecte des données

Cette section décrit le protocole mis en place pour la récolte de données à travers les différentes exploitations du projet. Les données collectées dans le cadre de ce projet sont des données météorologiques, des données de production, de pratique de gestions et des données de santé des sols. Dans un premier temps, le protocole général de collecte de données implémentées est décrit. La sous-section 5.2.2 présente l'acquisition des données de production. La collecte des pratiques de gestions est explicitée en 5.2.3 et la récolte des données caractéristiques de chaque champ est présentée en 5.2.4. L'acquisition des données météorologiques est décrite en sous-section 5.2.5.

5.2.1 Protocole général de collecte

Comme décrit dans la section 5.1, le réseau d'exploitation est très dispersé sur le territoire Canadien. Certaines variables collectées sont extraites à partir des paramètres biologiques des échantillons de luzerne récoltés. Les analyses chimiques des échantillons doivent être réalisées peu de temps après leur récolte. Certaines variables comme les données météorologiques doivent être acquises tout au long du cycle d'exploitation. Plusieurs stratégies sont adoptées pour la récolte. La stratégie générale est expliquée à la figure 5.1.

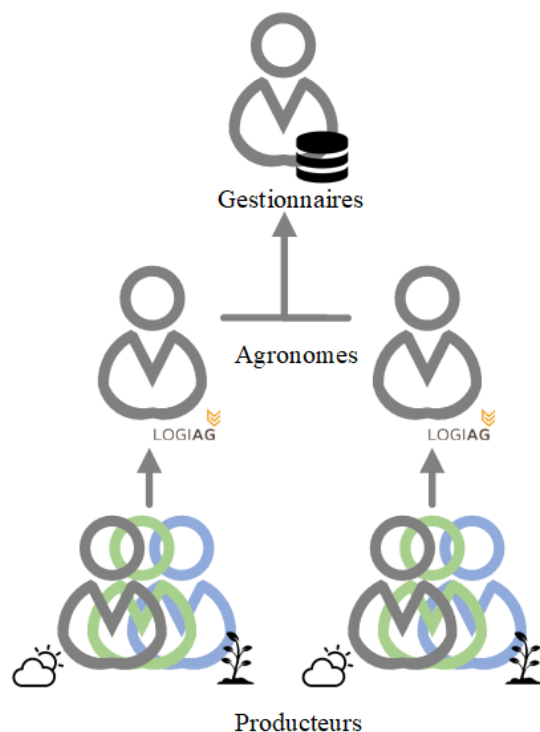


Figure 5.1 Organigramme du protocole de collecte des données

Les agronomes prenant part au projet sont déployés sur le territoire afin de pouvoir gérer la collecte des données sur l'ensemble du réseau de producteurs. Chaque agronome est responsable du suivi de 4 à 8 producteurs en fonction de leur proximité géographique. Leur rôle est de choisir les points GPS adéquats pour le positionnement des bornes de prélèvement, de s'assurer du bon prélèvement des échantillons de luzernes lors des récoltes, d'extraire les échantillons de sols et de mettre en place les stations météorologiques pour chaque ferme. Les agronomes sont aussi les points de contact avec les producteurs et permettent d'obtenir des retours sur les besoins des producteurs.

En fonction du type de données collectées, différentes stratégies sont adoptées. Ces stratégies sont expliquées dans les sections suivantes. Pour une grande partie des données, la collecte est effectuée par le biais d'une interface d'acquisition présentée en figure. 5.2

Figure 5.2 Interface de collecte des données

Le rôle de cette interface est d'acquérir les données nécessitant une entrée manuelle. Cet outil permet de standardiser et de limiter l'entrée de certaines données participants ainsi au pré nettoyage des données.

5.2.2 Collecte des données de production

Les données de production représentent les rendements obtenus lors de chaque coupe et la qualité de la luzerne récoltée. L'acquisition de ces données doit donc être effectuée 3 à 5 fois par saison. La récolte de ce type de données étant ponctuelle, il est choisi d'utiliser des méthodes de collecte manuelle. Le protocole utilisé pour la collecte des données de production est présenté en figure 5.3.

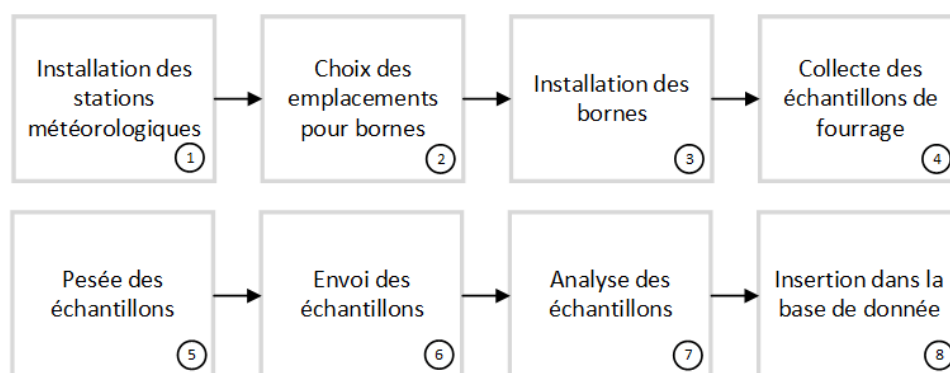


Figure 5.3 Protocole de collecte de données de production

Dans un premier temps, 12 balises GPS sont installées dans chaque champ de chaque exploitation. Ces balises GPS sont placées dans le champ à partir d'une étude de la productivité des

champs et d'images satellites. Elles représentent les points de prélèvement à partir desquels les échantillons sont prélevés. Afin de représenter toutes les caractéristiques de performances des champs, des points aux faibles, moyens et grands rendements sont disposés sur un même champ. Une fois positionnées, ces balises sont insérées dans le sol et sont utilisées comme références. Les résultats des analyses de ces points permettent ainsi d'extrapoler les performances moyennes de chaque champ.

Le déroulement du protocole de collecte s'effectue alors avant chaque coupe (maximum 5 jours avant). Un agronome est déployé chez le producteur pour effectuer le prélèvement d'échantillons de fourrages. Ces échantillons sont pesés sur place afin d'acquérir la quantité de fourrage frais produit sur une surface de 1 pied carré. Ces derniers sont ensuite congelés pour leur conservation et placés dans un contenant isotherme. Les échantillons sont enfin envoyés dans les 48h dans un laboratoire où des analyses de fourrages sont effectuées.

L'analyse de ces échantillons permet de produire une analyse de fourrage ainsi qu'un poids des fourrages séchés. Le poids sec et frais des échantillons est ainsi entré dans la base de données CC+ à l'aide d'une application développée par LOGIAG. Chaque échantillon de fourrage est ensuite analysé afin d'évaluer la qualité nutritive de la luzerne produite. Les variables utilisées pour décrire la qualité de la luzerne seront présentées plus loin, dans le tableau 5.1.

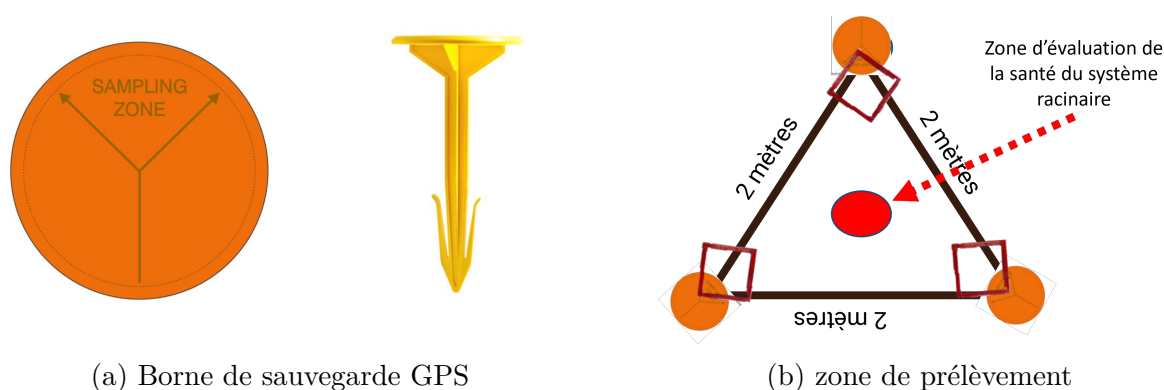


Figure 5.4 Outils de collecte des données

Une fois les données prélevées, elles sont ensuite entrées sous forme de rapport pour chaque borne dans CC+. Les données de rendement frais sont directement entrées à l'aide d'un moniteur juste après l'étape de pesée. Les analyses de fourrages sont entrées dans la base de données une fois effectuées. Ces données sont ensuite extraites de CC+ par le script python Query.py de la plateforme DPMS. La figure 5.5 présente l'interface d'entrée des mesures de

rendements dans l'application de collecte de données.

Figure 5.5 Interface de collecte des données de rendement

5.2.3 Collecte des données de pratiques de gestion

Les pratiques de gestions rassemblent toutes les activités effectuées sur le champ visant à l'augmentation des rendements et à la qualité de production. Ces activités peuvent être divisées en 3 sous-activités: Les pratiques de fertilisation organique, de fertilisation minérales (date et quantité appliquée) et l'identification des nuisibles et maladies. La figure 5.6 présente le protocole de collecte des données de pratique de gestion permettant de collecter ces trois sous-activités.

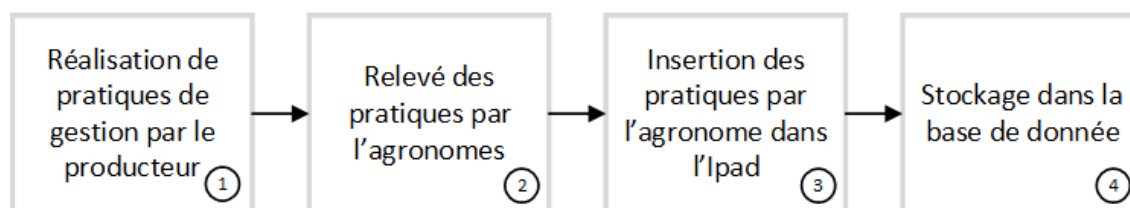


Figure 5.6 Protocole de collecte des données de pratique de gestion

Les pratiques de gestion rassemblent l'application de fumiers, d'azote, de calcium, de phosphore et de magnésium. Ces données sont collectées par les agronomes lors de leurs visites chez les exploitants et lors de chaque modification sur le champ. En général, les activités de fertilisation minérales sont effectuées en fin de saison et les activités de fertilisation organiques sont réalisées quelques jours suivant les récoltes. La date de ces activités ainsi que la quantité de produits appliquée sont enregistrées puis intégrées à la plateforme de données

du partenaire industriel Logiag: R+. Ces données sont ensuite extraites par l'entremise du script Query.py et implémentées dans DPMS.

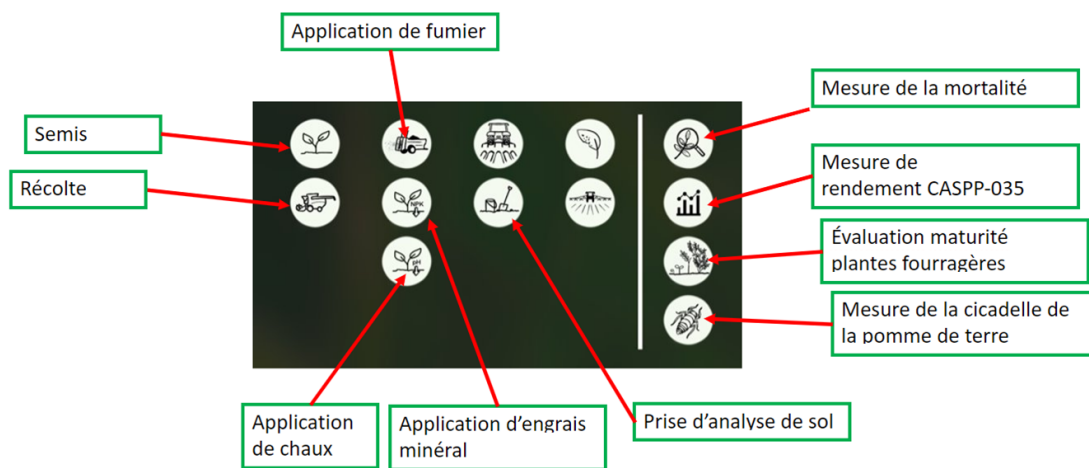


Figure 5.7 Interface de collecte de pratique de gestion

La figure 5.7 présente l'interface utilisateur de l'application de collecte permettant de récolter les données de pratiques de gestion.

5.2.4 Collecte des données de caractéristiques des sols

Afin d'évaluer la santé des sols de chaque champ, des relevés d'échantillons de sol sont effectués par des agronomes tous les 5 ans. Les données collectées proviennent de ces rapports ainsi que de nouveaux rapports effectués pour les besoins du projet. Ces rapports permettent de présenter le taux de nutriments, l'évolution de la structure des sols et ainsi d'identifier de potentielles carences pouvant limiter les rendements de productions. Les données de caractéristiques des champs rassemblent toutes ces informations relatives à la santé du sol. Ainsi, le type de sol, la texture, la quantité de minéraux ainsi que la quantité de matière organique présente dans le sol sont prélevés.

Les différents types de minéraux identifiables dans le cadre de ce projet sont de l'azote, du phosphore, du potassium, du magnésium et du calcium. Ensuite, l'acidité des sols est aussi acquise à partir du pH relevé dans les champs étudiés. L'ensemble de ces paramètres est prélevé en plusieurs points dans le champ permettant ainsi d'obtenir une valeur moyenne représentative de ces paramètres pour chaque champ. Ces données, une fois récoltées par les agronomes, sont ensuite transmises à la base de données R+ sous forme de rapport. Une extraction de ces données est enfin réalisée par le script Query.py.

5.2.5 Collecte des données météorologiques

Lors de la période de croissance et entre 2 récoltes, la qualité et les rendements de production peuvent être modifiés par les paramètres météorologiques. Ces paramètres sont en constante évolution et impact le développement de la luzerne tout au long du cycle d'exploitation. Les paramètres principaux ayant de l'influence sur le bon développement de cette légumineuse sont la température, la précipitation, les chutes de neige, la pression atmosphérique et les radiations solaires.

Afin de mesurer l'impact de ces variables météorologiques sur le cycle de croissance, un protocole de collecte automatisé est disposé. Ce protocole consiste en l'installation de stations météorologiques pour chaque ferme. D'autres paramètres sont acquis à partir de base de données externes. La figure 5.8 présente le réseau de stations météorologiques.

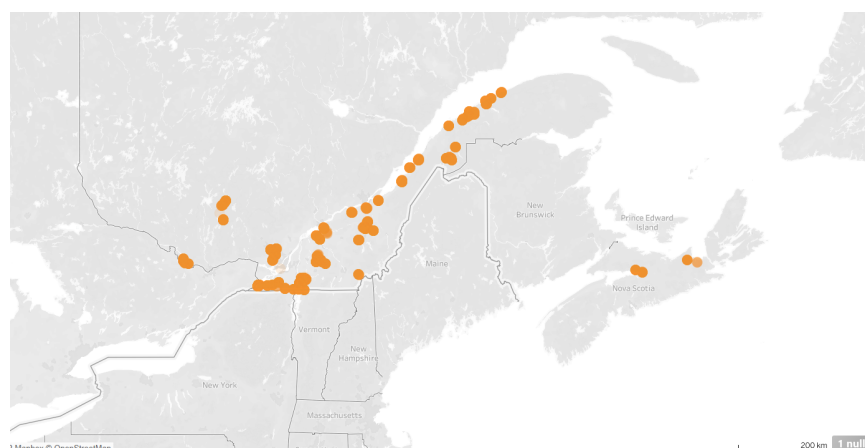
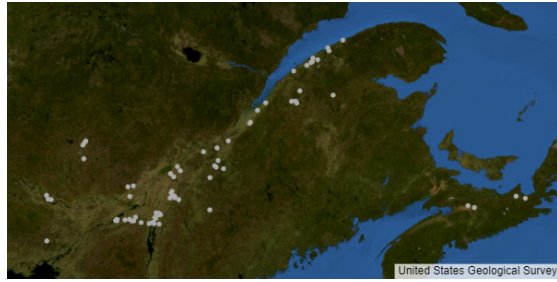


Figure 5.8 Répartition des stations météorologiques du projet

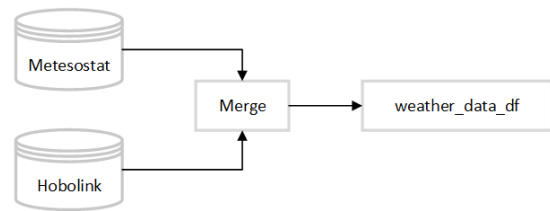
Afin d'augmenter la récolte des données météorologiques tout en adaptant au contexte de chaque partenaire industriel, l'extraction de ces données est réalisée à l'aide d'un script python et de l'API météoestat. L'API Météostat est un service public permettant d'extraire des données météorologiques sur une période spécifiée à partir d'un point GPS. Pour ce faire, les localisations de chaque ferme sont estimées par rapport à la moyenne géographique de l'ensemble des mesures prises sur chaque ferme. Ensuite, ces coordonnées GPS sont introduites dans l'API sur la période de 1er janvier 2021 jusqu'au 1er janvier 2022.

Les données récoltées sont des données de précipitations, de radiations solaires, de températures minimales, de température maximale et de pressions atmosphériques. Ces données sont agrégées par jour. Afin de vérifier la viabilité des données, un code python est adapté pour que seules les données des stations à moins de 50km de chaque ferme soient extraites. Le

script permettant d'effectuer ce protocole est situé dans une section du fichier `query.py`.



(a) Représentation géographique du réseau de stations Hobolink



(b) Protocole de collecte des données météorologiques

Figure 5.9 Protocole de collecte des données météorologiques

L'étape de collecte des données météorologiques permet d'obtenir une table de données contenant les principales variables environnementales impactant le cycle de croissance. L'ensemble des paramètres récoltés est résumé dans le tableau 5.1.

Table 5.1 Paramètres collectés dans le cadre du projet

Données	Unité	Occurrence	Provenance	Collecte	Précision
Rendements					
Poids	kg	Coupe	CC+	Manuelle	Borne
Qualité	SA				
NDF	%	Coupe	R+	Manuelle	Champ
ADF	%	Coupe	R+	Manuelle	Champ
Matière organique	%	Coupe	R+	Manuelle	Champ
Analyse des sols					
Azote	kg/Ha	Année	R+	Manuelle	Champ
Phosphore	kg/Ha	Année	R+	Manuelle	Champ
Potassium	kg/Ha	Année	R+	Manuelle	Champ
Calcium	ppm	Année	R+	Manuelle	Champ
Magnésium	ppm	Année	R+	Manuelle	Champ
Pratique de gestion					
Fertilisation organique	kg/Ha	Jour	R+	Manuelle	Champ
Fertilisation N-P-K-S	kg/Ha	Jour	R+	Manuelle	Champ
Météo					
Précipitation	mm	Jour	Hobolink	automatique	ferme
Température	Celsius	Jour	Hobolink	automatique	ferme
Pression	Pa	Jour	Hobolink	automatique	ferme

Le tableau 5.1 présente les variables collectées dans le cadre de ce projet. L'occurrence à laquelle ces attributs sont collectés est décrite dans ce tableau ainsi que la source primaire de ces données. La précision de collecte des données ainsi que le type de collecte de données y sont introduits.

L'étape de collecte des données permet d'obtenir plusieurs sources de données décrivant l'ensemble des paramètres déterminés lors de l'analyse du contexte 5.2 et ayant une influence sur le cycle de croissance de la luzerne. Ce protocole de collecte est effectué la première année et est réajusté en fonction des retours des producteurs lors des années suivantes. La prochaine étape consiste alors à structurer les données collectées dans une base de données ainsi que de préparer ces données. L'étape de structure et de préparation des données est décrite dans la section 5.3.

5.3 Préparation des données

La préparation des données dans le cadre de ce projet se déroule en 3 sous-étapes. Tout d'abord, une architecture de la base de données est présentée 5.3.1. Ensuite, les processus de nettoyages de variables collectées en section 5.2 sont décrits 5.3.3. Enfin, le formatage des tables de données est réalisé 5.3.4. Cette étape de préparation des données vise à fournir plusieurs tables de données formatées pouvant servir de base solide aux étapes de valorisation des données.

5.3.1 Structure de la base de données

L'étape suivante au processus de collecte de données est la préparation des données. Afin de réaliser cette étape, une base de données est structurée dans le but de stocker les changements effectués au cours des étapes de récoltes de données.

Les données agronomiques de rendements et de pratique de gestion sont dans un premier temps structurées à l'aide de la plateforme Jira permettant de produire des rapports structurés en fonction des besoins des scientifiques de données qui utiliseront cette plateforme comme source de vérité. Les données caractéristiques à chaque champ structuré dans la plateforme du partenaire industriel LOGIAG qui permet de produire 2 rapports structurés: (anciennes récoltes et analyse des sols). Aucune action n'est prise sur les données météorologiques puisqu'un pipeline de donnée est déjà structuré lors de la collecte des données. Cette étape est illustrée en figure 5.10. L'objectif des sous-sections suivantes est donc de bâtir l'architecture de code du module DPMS.

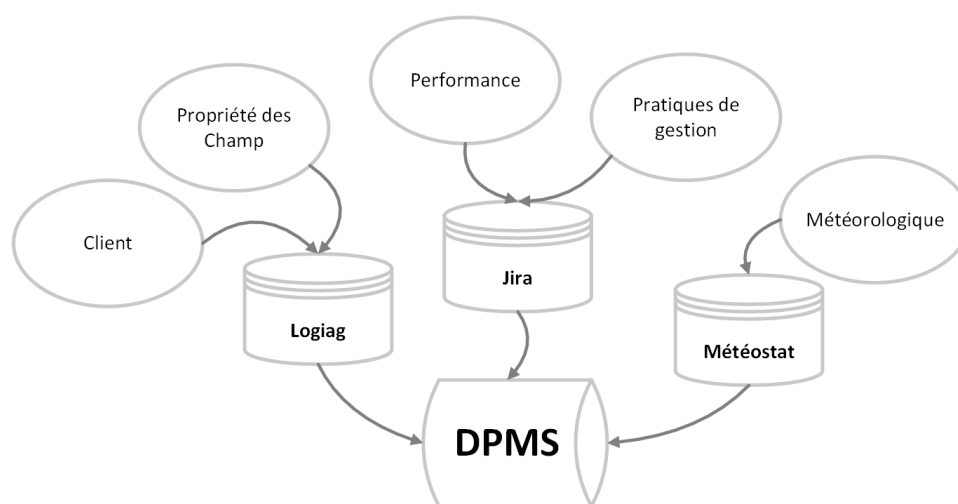


Figure 5.10 Structure de la base de données

À partir de ces 3 sources de vérité, les données sont dans un premier temps importées sous forme de rapport. L'architecture est cependant conçue pour accueillir plus tard des connexions de données en temps réel. Il est possible de débiter les étapes de nettoyage et de préparation des données. La figure 5.11 présente l'architecture proposée dans la méthodologie et appliquée au cas d'étude.

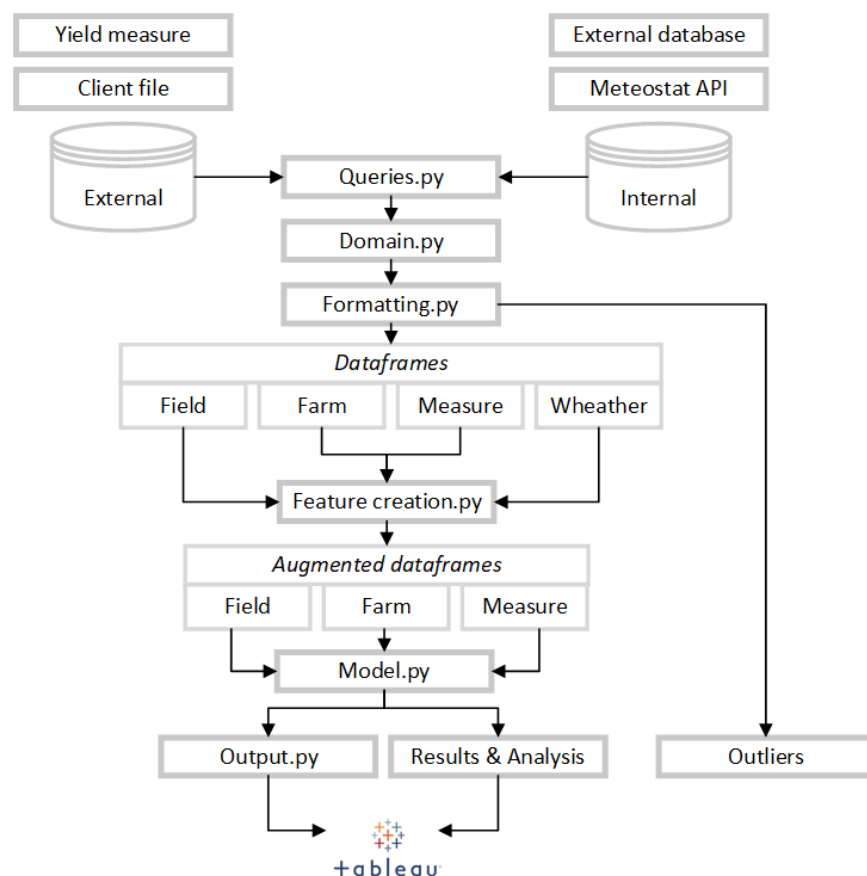


Figure 5.11 Architecture de la méthodologie

Le script Query.py permet l'extraction des données présentes dans les différentes sources de données. Ce dernier communique les tables brutes de données au fichier Domain.py. L'utilité de ce dernier est de renommer toutes les variables sous un format standard afin de renforcer la résilience du code construit. Ensuite, l'objectif du fichier Formatting.py est de construire 4 tables de données nettoyées et de fournir ces tables au fichier Feature_creation.py. Ce fichier a aussi pour objectif d'extraire les données aberrantes pour de futures analyses. Ensuite, le fichier Feature_creation.py débute le processus de valorisation des données en calculant de nouveaux indicateurs basés sur les données existantes.

Des modèles peuvent ensuite être construits à partir de ces tables de données augmentées puis

être extraient pour les futures étapes de visualisation dans l’outil tableau. Les sous-sections suivantes présentent ce processus en détail.

5.3.2 Nettoyage des noms de colonnes

Afin de standardiser les noms des attributs de colonnes pour chaque table de données collectées, les noms de colonnes sont modifiés en utilisant le format suivant.

Tout d’abord, le projet étant une collaboration entre plusieurs universités canadiennes, la langue choisie pour le format de la base de données est l’anglais. Les attributs des colonnes sont donc traduits en anglais. Ensuite, les espaces sont remplacés par le caractère ‘_’ et chaque variable est renommée suivant le format: `format_recency_variable`. Ainsi chaque variable indique clairement ce qu’elle mesure, sa temporalité et son domaine d’application. Par exemple, la variable mesurant initialement le poids de chaque échantillon collecté était nommée ‘Poids (kg)’ et devient sous le nouveau standard ‘`format_weight_kg_borne`’.

L’ensemble des noms de colonnes est ainsi reformaté dans le fichier `Domain.py` afin d’augmenter la résilience du code face à de futurs changements et d’améliorer la compréhension globale du code source de la plateforme développée.

5.3.3 Nettoyages des données

À la suite de la structuration de la base de données, plusieurs opérations doivent être effectuées afin de nettoyer les données structurées. Les données aberrantes doivent dans un premier temps être identifiées puis modifiées ou supprimées. De plus, les données manquantes doivent être collectées à nouveau ou remplacées par des données statistiques. Enfin, il est nécessaire de vérifier l’existence de duplicata dans la base de données. Si des duplicata existent, une stratégie doit alors être mise en place pour dissocier les enregistrements à garder de ceux à supprimer. Il est possible de retrouver la structure du nettoyage des données au sein du script python `Formatting.py`.

Données aberrantes

Dans un premier temps, une recherche des données aberrantes est appliquée aux données de production, au positionnement des balises de mesure ainsi qu’aux données de pratiques de gestion. Les attributs traités sont les rendements de productions, le nombre de coupes effectué, la date de coupe ainsi que la quantité d’intrants appliquée sur les cultures. Ces attributs sont présentés au tableau 5.2.

Table 5.2 Détection de valeurs aberrantes

Attribut	Limite inférieure	Limite supérieure	% aberrant
format_poids_kg_borne	0.1 kg	1 kg	0.5%
format_cut_date	25 avril	15 novembre	0.1%
format_measure_latitude	41.605601	50.243743	0.04%
format_measure_longitude	-122.202577	-66.097589	0.04%
format_field_mineral_fertilization	0kg/h	1000kg/h	2%
format_field_organic_fertilization	0kg/h	200 TM/h	1%
format_field_manure_application	5kg/h	15kg/h	0.4%
format_farm_cut	1	5	0%

La détection des données aberrantes est tout d'abord effectuée en se basant sur la méthode du Zscore. En effet, cette technique est privilégiée pour faciliter l'implantation et, car elle est très efficace pour détecter les points très distants des valeurs normales. La collecte des données étant réalisée manuellement, les causes possibles des valeurs aberrantes sont des erreurs typographiques lors de l'entrée, des erreurs de format entre le système de syntaxe anglo-saxon et le système international (, et .) et des erreurs d'unités. Dans tous les cas, ces erreurs aboutissent à des valeurs aberrantes très distantes des valeurs moyennes et sont facilement repérables avec le Zscore. Ainsi, il est décidé d'utiliser le critère $\sigma \leq 2$ pour chaque attribut présenté dans le tableau 5.2. La figure 5.12 présente une visualisation des valeurs aberrantes pour les données de rendement de production.

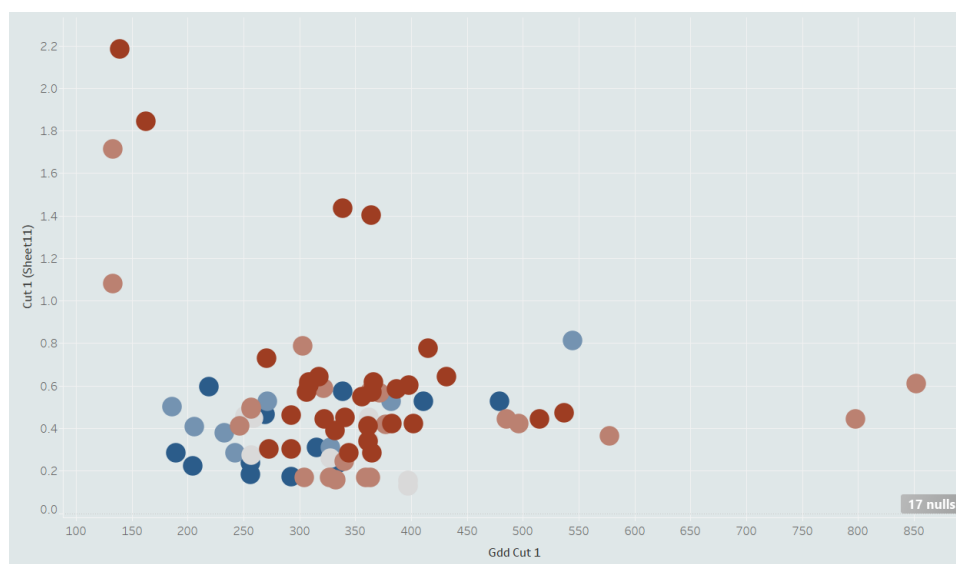


Figure 5.12 Visualisation des valeurs de rendements aberrants

À la suite de cette première étape d'identification, les résultats de l'analyse des valeurs aberrantes sont présentés et confrontés à un comité d'experts du domaine. Des filtres fixes sont ensuite créés afin de connaître les limites d'acceptabilités possibles des entrées et de les intégrer directement à l'application de collecte de données. Ces filtres sont présentés dans le tableau 5.2. En parallèle, les données aberrantes identifiées sont isolées et stockées dans un fichier 'Outliers'. Les agronomes et producteurs associés à ces valeurs sont alors contactés afin de corriger ces erreurs.

Données manquantes

La collecte des données pour l'année 2021 est effectuée sur la période de mars à novembre. À la fin de la période de collecte, les enregistrements manquants sont identifiés. Le traitement des données manquantes est réalisé en deux étapes. Tout d'abord, les producteurs associés aux enregistrements manquants sont contactés afin d'assurer la cause de ces manquements et qu'ils ne sont pas dus à des oublis lors de l'entrée dans la base de données. Ensuite, les colonnes concernées sont identifiées et traitées en fonction de leur contexte. Ainsi, les attributs concernés par le traitement de données manquantes sont les données de production, les données de pratiques de gestion et les données météorologiques. Ces attributs sont présentés dans le tableau 5.3.

Table 5.3 Traitement des valeurs manquantes

Attribut	% manquantes
format_poids_kg_borne	1.2%
format_cut_date	0%
format_field_mineral_fertilization	3.5%
format_field_organic_fertilization	1.5%
format_daily_farm_temp_avg	0%
format_daily_farm_temp_min	0.4%
format_daily_farm_temp_max	0%
format_daily_farm_temp_precip	7.9%
format_daily_farm_temp_wdir	73.3%
format_daily_farm_temp_wspd	72,1%
format_daily_farm_temp_wpgt	85.8%
format_daily_farm_temp_pres	2.0%
format_daily_farm_temp_tsun	45.3%
format_daily_farm_temp_GDD	0%
format_farm_cut	0%

Les données météorologiques proviennent en priorité des données de Météostats. Lorsque des enregistrements sont manquants, les données disponibles dans le réseau Hobolink sont consultées et intégrées à l'ensemble de données. Si les données ne sont pas présentes dans le réseau Hobolink, la valeur de l'enregistrement est remplacée par celle de la station la plus proche contenant des données.

Certains attributs de données météorologiques possèdent un fort taux de valeurs manquantes. Les radiations solaires ainsi que les informations relatives aux vents sont très peu disponibles sur le réseau de station sondée. Il a été choisi de le spécifier dans ce mémoire, car ces attributs peuvent influencer les rendements de production de la luzerne. Ces attributs devront être récoltés dans les prochaines années du projet.

Pour les données agronomiques, l'identification des agronomes associées aux valeurs manquantes permet la réalisation de sondages afin de récupérer les données manquantes. Pour les prochaines années, l'affichage des données manquantes pour que l'utilisateur soit averti sera utilisé. Enfin, lors de l'analyse des données manquantes, il a été remarqué que certains attributs n'avaient pas été collectés. Ces attributs tels que le numéro de coupe à laquelle la récolte est réalisée ont été collectés et nettoyés à la suite de cette étape. Des sondages ont été bâtis pour extraire ces informations chez les agronomes.

Données dupliquées

La base de données contenant les données de production est basée sur l'évènement. Ainsi, chaque enregistrement correspond à la modification d'un état des bornes de mesure. Ainsi, un premier tri est effectué afin de sélectionner uniquement les enregistrements d'intérêt.

5.3.4 Formatage des données

L'étape de nettoyage des données est suivie par le formatage. L'objectif de cette sous-étape est de créer un nombre limité de tables de données structurées et standardisées qui serviront à répondre aux besoins du projet. Pour ce faire, 5 tables sont produites. Ces tables de données sont décrites dans le tableau. 5.4

Table 5.4 Formatage des données en tables agrégées

Tables	Description	Dimension
Data_Luzerne_Farm_Formatted	Table contenant les informations agrégées au niveau de la ferme, tel que le rendement de la ferme annuel, le rendement par coupe ainsi que des informations moyennes sur les fermes voisines.	13 x 201
Data_Luzerne_Field_Formatted	Table contenant les informations agrégées au niveau du champ, tel que les rendements de production, la santé des champs ainsi que les analyses des pratiques de gestion	13 x 4257
Data_Luzerne_Measure_Formatted	Table contenant toutes les mesures initiales de rendements agrégées au niveau de la borne	25 x 7842
Big_Wheather_data_Raw_Formatted	Table contenant toutes les mesures des stations météorologiques, chaque enregistrement représente un jour pour une station, il est constitué pour une année de 28 259 enregistrements.	17 x 28259
Client_account_data_Formatted	Table de données représentant les informations des exploitants qui utilisent le service.	11 x 79

La figure 5.13 présente l'architecture des variables telles que formatées dans le cadre de ce projet. Les 5 tables de données y sont représentées ainsi que les liens de communication qui les regroupent.

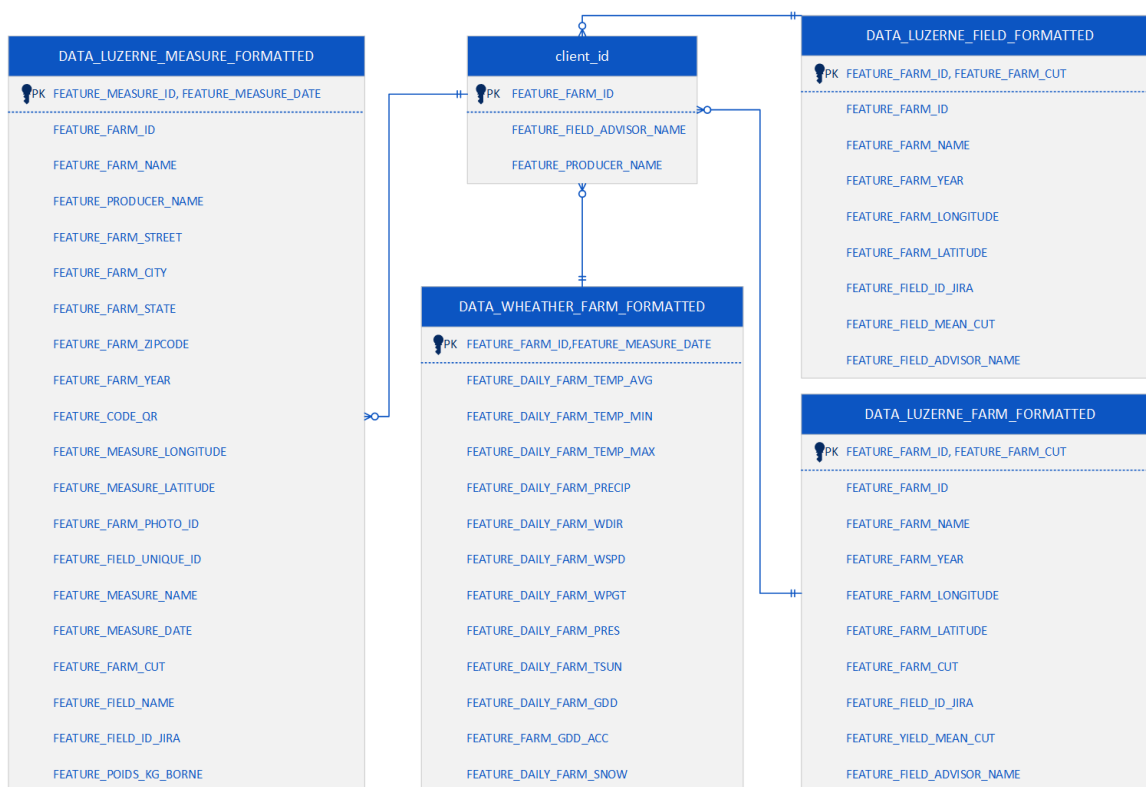


Figure 5.13 Structure de la base de données formatées

La création de ces 5 tables de données permet d'établir une source de connaissance solide ainsi que de valider l'interprétation facile de l'ensemble des données collectées. Cette étape valide aussi la préparation des données. À la fin de cette étape, 51 variables sont formatées et préparées. Ces variables sont ensuite utilisées comme base permettant de générer de nouveaux attributs.

5.4 Analyse et augmentation de la base de données

L'analyse et l'augmentation de la base de données ont pour objectif de générer de nouveaux attributs à partir des attributs formatés dans la section 5.3.4. Ainsi, pour chaque table de données, de nouveaux attributs sont générés. Ces attributs sont représentés à la figure 5.14.

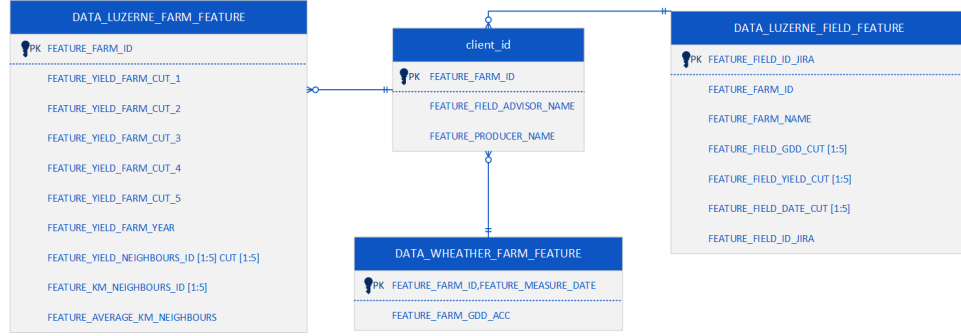


Figure 5.14 Structure des variables générées à partir des données formatées

Au niveau de la ferme, les attributs créés à partir des données formatées sont l'estimation des rendements par coupe, l'estimation du rendement total de la ferme par an, le rendement moyen des 5 fermes les plus proches, la distance moyenne des 5 fermes les plus proches. Ces attributs permettent donc de connaître les performances de production annuelle d'une exploitation ainsi que de les situer par rapport aux exploitations voisines. De plus, les données météorologiques sont augmentées par la production de deux nouveaux indicateurs agronomiques: les taux de degrés jours quotidiens et l'accumulation des degrés jours à partir du 1er avril. Ainsi, la notion de degré jour (GDD) est définie dans ce contexte industriel comme le nombre de degrés croissant supérieur à 5 degrés Celsius à chaque jour. La formule suivante est utilisée comme définition des degrés jours:

$$GDD_3 = \frac{T_{max} - T_{min}}{2} - 5 \quad (5.1)$$

D'après le contexte industriel, l'accumulation des GDD est un bon indicateur pour prévoir les dates de coupes. Ainsi, la somme cumulative de l'attribut GDD_5 est enregistrée dans un autre attribut nommé `Feature_farm_gdd_acc`. Au niveau de chaque champ, les attributs décrivant le nombre de degrés jour au moment de chaque coupe, le rendement estimé de chaque champ par coupe ainsi que la date de coupe réelle sont ajoutés à la table de données. L'ajout de ces attributs finalise l'augmentation de la base de données et permet de préparer les futures étapes de visualisation. L'ensemble des manipulations décrites dans cette section est réalisé par l'intermédiaire du script `feature_creation.py`.

5.5 Outils de visualisation des données

À partir de l'étape de préparation effectuée en section 5.4, les données obtenues sont structurées, nettoyées et augmentées par la création d'indicateurs de performances. La base de

données obtenue peut être interprétée par des outils de visualisation. Les données sont mises à disposition dans la plateforme tableau afin de produire un outil d'aide à la décision capable de communiquer de nouvelles connaissances aux producteurs.

Dans un premier temps, la visualisation des données est séparée en plusieurs tableaux de visualisation basés sur la compréhension des besoins du marché ainsi que sur les différents niveaux d'agrégations des données disponibles. Cette section présente les outils utilisés pour la visualisation des pratiques de gestion 5.5.1. Ensuite, la visualisation des rendements de production est présentée en sous-section 5.5.2. D'autres outils de visualisation permettant d'établir la santé de chaque champ sont alors représentés 5.5.1 . Enfin, des tableaux de visualisation du réseau général sont réalisés. Ces outils permettent ainsi aux gestionnaires d'établir la répartition des performances globales du réseau d'exploitants.

5.5.1 Visualisation des pratiques de gestion

La spécificité du secteur fourrager étant la persistance des champs exploités, les pratiques de gestions effectuées par le producteur au cours des 4 années d'exploitation ont un impact significatif sur le rendement et la qualité de production de la luzerne. La connaissance de l'ensemble des manipulations effectuées sur un champ apporte une information supplémentaire au producteur et permet de mieux comprendre l'impact de ces pratiques sur ces performances de production.

L'outil de visualisation des pratiques de gestions affiche pour chaque producteur, toutes les manipulations effectuées pour chaque champ. Les différentes pratiques présentées sont: l'application de fumier, l'application de manure, les dates de coupes et les pratiques de fertilisation minérales. Cet outil ajoute aussi les pratiques moyennes des 5 fermes voisines les plus proches, permettant ainsi au producteur de comparer ses pratiques à celles de ses voisins. Cet outil affiche les dates de chaque action et spécifie les quantités d'application. Enfin, le but de cet outil est de recommander des dates de coupes potentielles lorsqu'une quantité suffisante de données sera accumulée. Les principaux indicateurs mis en place sont présentés en figure 5.15 puis sont ensuite expliqués.



Figure 5.15 Représentation des pratiques de gestion

Carte et représentation des champs

Afin d'assurer aux producteurs que le champ qu'il consulte est le bon, une carte du champ sondé est affichée avec les bornes actuellement en fonctionnement. Cette carte permettra à plus long terme d'afficher des informations supplémentaires concernant les pratiques de gestion réalisées. De plus, le producteur peut connaître l'emplacement GPS des bornes situées dans son champ.

Ligne du temps

L'outil de ligne du temps permet d'afficher les différentes pratiques de gestion des agriculteurs. Ainsi, au cours d'une saison d'exploitation, le producteur effectue des pratiques de fertilisation minérale, de fertilisation organique, des étapes de semis et de récoltes. Chaque opération corrélée avec la composante climatique exerce une influence sur les rendements et la qualité de production. La date, la quantité et le type de ces opérations sont donc collectés et affichés dans ce premier outil de visualisation. Ces données pourront ensuite être comparées par région afin de mettre en évidence des bonnes et de mauvaises pratiques. Ce premier outil est présenté en figure 5.15.

5.5.2 Visualisation des rendements de production

Afin de permettre aux producteurs de suivre les performances de production de leur exploitation, l'outil de visualisation des rendements de production présente pour chaque année d'exploitation le rendement estimé à partir des échantillons collectés lors de chaque récolte. La figure 5.16 présente cet outil de visualisation. Ce dernier est constitué d'une carte de champ ainsi que d'un graphique présentant les rendements par coupe.

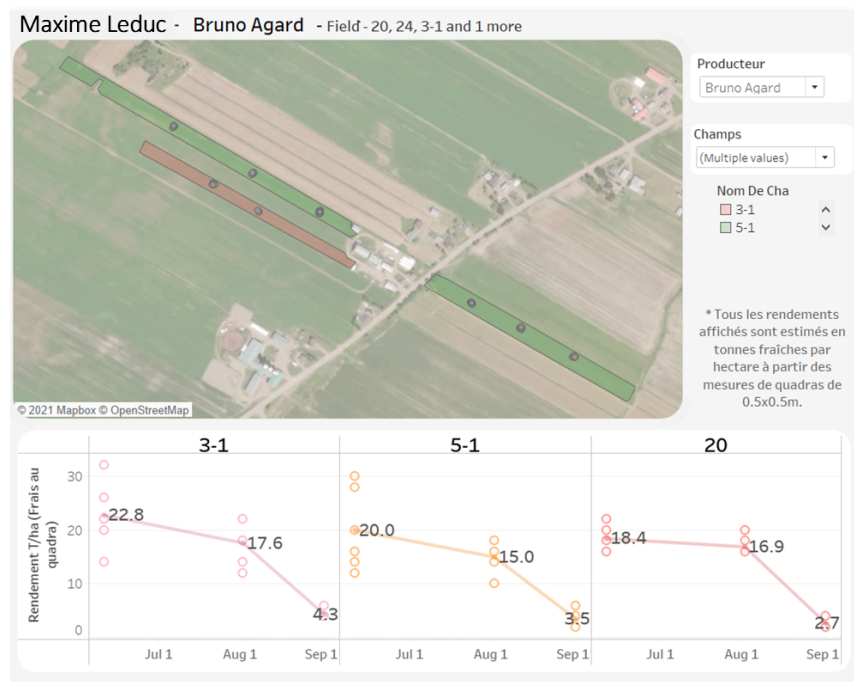


Figure 5.16 Représentation des rendements de production

Carte des champs

L'outil de carte de champs permet aux producteurs d'identifier les bornes de mesures ainsi que les performances globales de ces champs en termes de rendement de production.

Rendements par coupes

Ce graphique présente une estimation du rendement en T/H de fourrage frais récolté. Il indique la date à laquelle chaque champ est récolté, le nombre de coupes effectuées sur chaque champ ainsi que la répartition des rendements collectés à chaque point de mesure. Il est possible d'identifier les parties du champ ayant de plus faibles rendements de productions. Ce graphique présente le rendement moyen des 5 exploitations les plus proches permettant

ainsi aux producteurs de situer ses performances de productions. Plus précisément, cet outil est présenté en figure 5.16. Les perspectives possibles d'améliorations de cet outil sont la prévision des rendements des prochaines coupes ainsi que de la qualité de production reposant sur les analyses de fourrages précédentes.

5.5.3 Visualisation de la santé des champs

L'outil de visualisation de la santé des champs permet aux producteurs d'acquérir une vision globale de la qualité des sols des champs sondés. Il est constitué d'une carte des champs indiquant la santé générale des champs et d'un résumé de l'analyse des sols identifiant les carences potentielles de chaque champ.

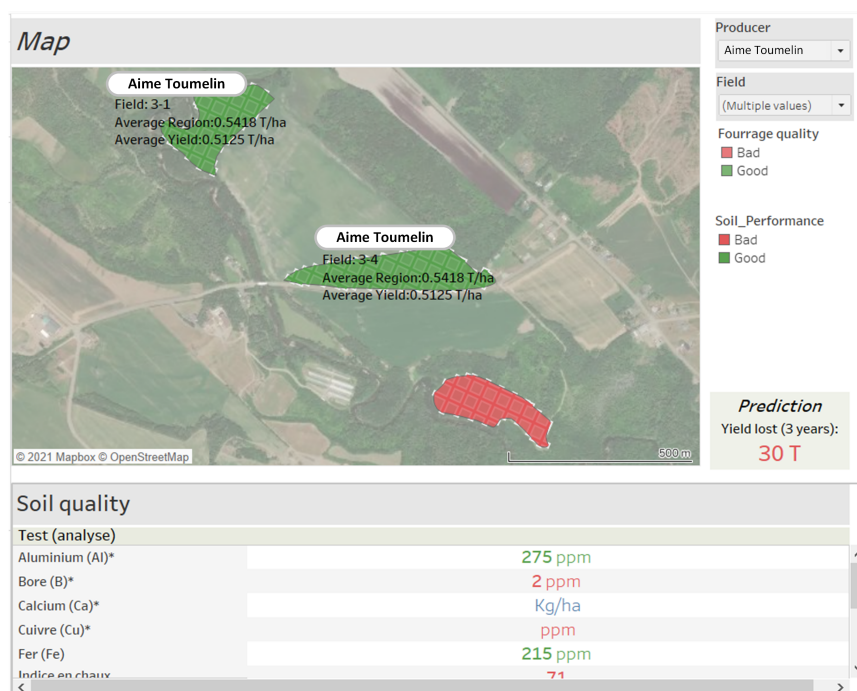


Figure 5.17 Représentation de la santé des champs

Carte des champs

La carte des champs expose un état général de la santé de chaque champ suivant un code couleur divisé en trois états. Elle permet aux producteurs d'identifier ses champs et de consulter l'analyse des sols.

Analyse de la santé des sols

Cet outil présente les composantes générales de l'analyse des sols pour un champ. Des paramètres tels que le taux de minéraux dans le sol, la structure organique du substrat de croissance et l'humidité des sols y sont présentés. La liste complète des paramètres fournis dans l'analyse des sols est présentée au tableau 5.5.

Chaque paramètre est coloré par un indicateur permettant d'identifier un surplus ou une carence par rapport à la valeur souhaitée. Les données utilisées pour analyser ces paramètres sont tirées des analyses de sols de chaque champ effectué au minimum tous les 5 ans ainsi que des analyses de fourrages. Ces données sont affichées par l'intermédiaire d'une carte donnant accès à l'analyse de sols du champ, aux dernières analyses de fourrages ainsi qu'à une estimation de ces résultats. Ainsi des indicateurs de couleurs graphiques sont affichés afin de synthétiser les résultats de ces analyses. Les couleurs de chaque champ indiquent donc la santé globale de ce dernier. Ce dernier outil est présenté en figure 5.18.

Table 5.5 Paramètres de l'analyse de sol

Attribut
format_organic_matter_field
format_ph_field
format_phtamp_field
format_potassium_field
format_phosphore_field
format_calcium_field
format_magnesium_field
format_ppm_cec_field
format_al_field
format_fertilization_N
format_fertilization_P
format_fertilization_K
format_fertilization_dose
format_manure_quantity
format_lime_quantity
format_lime_type
format_lime_humidity
format_crop_historic



Figure 5.18 Outils de présentation de la santé des sols

5.5.4 Visualisation générale du réseau d'exploitants

Enfin, un dernier outil de visualisation est conçu pour les gestionnaires de projet permettant de visualiser la santé globale du réseau d'exploitation. Ce tableau contient l'ensemble des données du réseau d'entreprises et doit donc être accessible uniquement aux gestionnaires de projet. Cet outil est identique aux outils précédemment présentés à l'exception qu'il ne possède aucune restriction sur les accès aux données des fermes.

Les outils présentés ci-dessus présentent donc des exemples d'indicateurs de performances permettant de présenter les performances individuelles de chaque partenaire. Ces indicateurs ont été déterminés en fonction du contexte d'application et peuvent être utilisés pour d'autres domaines d'études. Ainsi, une méthode similaire pourrait être appliquée pour comparer les données individuelles des PME aux données du réseau respectif. Ces exemples illustrent donc la faisabilité de cette méthode sans contraindre la confidentialité et la sécurité des données de chaque entreprise.

Les perspectives possibles de développement de cet outil sont l'utilisation de techniques d'apprentissages machine afin d'identifier et de déterminer les paramètres environnementaux exerçant une influence sur la dégradation des caractéristiques de chaque champ.

5.6 Optimisation des performances

La visualisation des performances actuelles de chaque PME étant présentées, l'étape suivante, réalisée dans le cadre de la méthodologie, consiste à l'amélioration des performances de

chaque ferme du réseau de PME. Au cours de cette étape et d'après la méthodologie, un outil de comparaison ainsi que des indicateurs de gestion sont mis en place. Pour ce faire, plusieurs indicateurs sont générés afin d'apporter des connaissances supplémentaires aux utilisateurs finaux de l'outil. Dans le cadre de l'optimisation des performances, 2 indicateurs déjà introduits dans ce mémoire sont affichés et permettent d'améliorer les performances de chaque PME. Ces deux indicateurs sont les degrés jours et l'outil de comparaison des performances de chaque ferme.

5.6.1 Outil de comparaison

À partir de la figure 4.7, l'analyse des flux d'informations est caractérisée dans ce contexte industriel par l'analyse des caractéristiques du sol présent au sein de chaque exploitation. Les flux d'informations sont par exemple, l'évolution des taux de minéraux dans le sol, le pourcentage de matière organique ou bien l'acidité des sols de chaque champ. De plus, les pratiques de gestion sont caractérisées dans ce contexte industriel par l'ensemble des modifications apportées par l'agriculteur sur le champ. Ces modifications se traduisent par une modification des paramètres de cultures ainsi que par l'introduction de ressources externes sur la parcelle exploitée.

Le suivi des performances de production se caractérise par les rendements obtenus en tonne par hectare pour chaque champ ainsi que par la qualité de la luzerne obtenue. La comparaison de ces composantes sur l'ensemble du réseau permet de dégager des entreprises plus performantes ainsi que des entreprises moins performantes. Afin de transmettre ce concept aux partenaires industriels, un outil de comparaison est alors conçu et présenté à chaque producteur. Cet outil place en perspective les flux d'informations, les pratiques de gestion et les performances de production de chaque producteur par rapport aux 5 exploitations les plus proches.

En fonction des similarités de chaque ferme et du contexte, chaque exploitant peut remettre en question ses pratiques de gestion en comparant ses performances avec celles de ses voisins. Cet outil constitue alors le premier indicateur mis en place permettant d'améliorer les performances de production des PME du réseau décentralisé. Les données de cet outil sont stockées dans une table de données nommée `data_luzerne_farm_feature` constitué de 51 attributs pour les 76 fermes du réseau étudié.

5.6.2 Degrés jours

Le second indicateur mis en place est le calcul personnalisé des degrés jours. Cet indicateur s'avère particulièrement utile pour suivre l'évolution de la plante au cours de son cycle d'exploitation. Comme la croissance de la plante est très corrélée à cet indicateur, des paliers fixes peuvent être déterminés et communiqués aux producteurs afin d'orienter leurs pratiques de gestion.

D'après la littérature, les experts du domaine préconisent d'effectuer la première coupe aux alentours de 150 degrés jours. L'intégration de cet indicateur à l'outil permet alors aux producteurs de suivre ces recommandations. La compilation de cet indicateur avec l'outil de comparaison permet de comprendre en fonction du contexte géographique à quelle accumulation de degrés la première coupe est la plus optimale. Cette réflexion peut donc être élargie aux autres pratiques de gestion effectuées par le producteur.

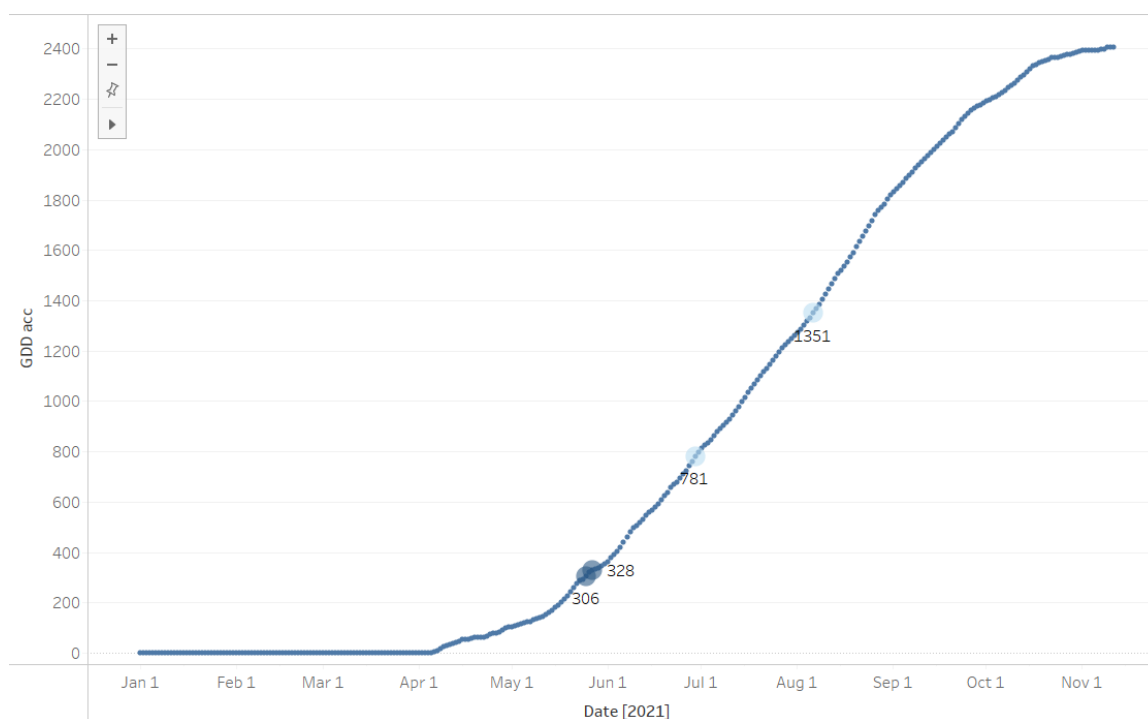


Figure 5.19 Courbe de Coupes en fonction des degrés jours d'une exploitation

La figure 5.19 permet pour un producteur de visualiser les différentes coupes de luzerne en fonction des degrés jours. Ainsi, pour cette exploitation, la première coupe est effectuée en deux parties à 306 et 328 degrés jours. La seconde coupe est effectuée à 781 degrés jours et la dernière coupe à 1351 degrés jours.

5.6.3 Perspectives d'améliorations

À partir des deux outils générés dans cette section d'amélioration, l'utilisation des techniques d'apprentissages supervisées et de segmentations peuvent permettre de créer des groupes de producteur similaires ainsi que de classifier leurs performances en catégories interprétables pour les producteurs et les agronomes. Ces classifications pourraient être affichées directement pour les producteurs par le biais de l'outil diagnostique. Le but principal de cette section est alors de dégager des recommandations sur les bonnes pratiques à réaliser en fonction du contexte de chaque entreprise présente dans le réseau. Ces voies de travail constituent alors des perspectives possibles d'amélioration pour le futur développement de l'outil.

5.7 Déploiement

Cette section décrit les différentes étapes nécessaires au déploiement de l'outil dans le réseau de PME. Ces étapes sont réalisées afin d'assurer l'anonymat des partenaires du réseau tout en affichant des informations pertinentes permettant d'améliorer les performances de chaque exploitant. L'étape de déploiement consiste dans un premier temps à l'anonymisation et la sécurisation des données. Dans un second temps, un processus de gestion des accès est réalisé. Enfin, une stratégie de maintenance et d'amélioration est mise en place.

5.7.1 Anonymisation des données

Afin de protéger les données des partenaires du réseau de PME, un processus d'anonymisation des données sensibles est effectué sur les données collectées. Tous les attributs permettant de relier les performances d'une exploitation à son propriétaire sont modifiés par un code python hébergé dans l'exécution du code principal. Les noms d'agriculteurs, les noms d'agronomes, les adresses des exploitations ainsi que les noms des champs sont remplacés par des identifiants. Les données relatives aux performances ainsi qu'aux pratiques des gestions de chaque exploitation ne sont communiquées en l'état qu'aux producteurs concernés.

Afin de produire des connaissances utiles avec l'outil de comparaison présenté en section 5.6.1 tout en assurant la protection des partenaires, tous les calculs sont effectués dans le code Python. Seuls des fichiers avec les variables déjà anonymisées sont alors introduits dans l'outil de visualisation. La figure 5.20 présente alors ce processus de protection des partenaires du réseau de PME.

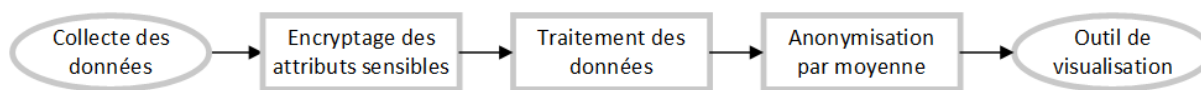


Figure 5.20 Processus de protection des données

Ce processus représente une première couche de sécurité et permet d'assurer la protection des données des producteurs lors du traitement des données. À la suite de cette première couche, une seconde couche de protection est réalisée par l'outil de visualisation et permet de limiter l'accès aux codes sources de la plateforme. Cette seconde protection est présentée dans la sous-section suivante.

5.7.2 Sécurisation des données

Lors de la transmission des données traitées entre le code source de la plateforme et l'outil de visualisation tableau, un processus de sécurisation des données doit être intégré afin qu'il soit impossible pour une institution malveillante d'extraire des informations présentes dans le code de l'application tel que les clés d'encryptions ou les rendements de production des autres fermes à partir de l'interface d'un producteur. Une architecture en valve doit être utilisée afin de permettre à chaque producteur de consulter ses propres informations sans avoir accès à celles des autres producteurs ainsi qu'au code source de la plateforme. L'application tableau est utilisée afin de garantir cette sécurité. La figure 5.21 illustre l'architecture choisie permettant d'assurer cette sécurité.



Figure 5.21 Architecture en valve choisie pour la visualisation des données

Cette sous-section permet donc d'assurer l'unilatéralité des flux de données ayant pour résultante la protection des données des utilisateurs. Une fois cette action réalisée, différents accès doivent être attribués en fonction des niveaux d'autorisations des utilisateurs pouvant être amenés à consulter cet outil. La section suivante présente donc les outils utilisés pour la gestion des accès et des niveaux d'accessibilité de la plateforme.

5.7.3 Gestion des accès

À partir de la compréhension du contexte industriel du réseau de PME étudié, plusieurs niveaux d'accès correspondant à différents niveaux d'utilisateurs sont générés. Cette sous-section a pour objectif de clarifier la notion d'accès à la plateforme ainsi que de définir les niveaux d'accès des utilisateurs. Dans le cadre de ce projet, 3 niveaux d'utilisateurs sont identifiés. Ces niveaux sont présentés à la figure 5.22.

	Niveaux d'accès		
	Administrateurs	Agronomes	Producteurs
• Réseau global	✓	✗	✗
• Producteurs affiliée	✓	✓	✗
• Exploitation unique	✓	✓	✓

Figure 5.22 Présentation de différents niveaux d'accès

Les utilisateurs possédant le plus d'accès sont les administrateurs. Ces derniers sont chargés du soutien de la plateforme, du soutien aux utilisateurs ainsi que de l'analyse des performances du réseau global. Leurs responsabilités nécessitent d'obtenir les accès du réseau général pour analyser les tendances globales des performances du réseau de PME.

Le niveau agronome permet de consulter toutes les informations d'un groupe d'exploitation. En règle générale, chaque agronome est responsable de différentes fermes. Ce niveau permet aux agronomes de consulter les données relatives à chaque ferme sous sa responsabilité. Il peut ainsi comparer les performances de ses fermes dans son groupe et recommander des pratiques pour ses producteurs.

Le niveau producteur permet à chaque producteur de consulter les performances relatives à sa propre exploitation. Ce niveau est le plus restrictif et le plus répandu dans le réseau de PME.

La gestion de ces trois niveaux d'accès est réalisée par l'intermédiaire de filtres dynamiques

dans Tableau. À partir de la table de données des clients de la plateforme, une table d'accès est générée. Cette table est hébergée sur les serveurs de Tableau et contient les noms d'identifiants, les clés d'accès tableau et les adresses emails de chaque utilisateur de la plateforme. À partir de cette table, des filtres dynamiques sont générés permettant de limiter les accès en fonction de l'utilisateur et de son niveau d'accès. La gestion des accès constitue la dernière étape avant le lancement en production de la plateforme d'aide à la décision. À la fin de cette étape, la plateforme est mise en ligne pour l'ensemble des utilisateurs et un plan de maintenance et d'amélioration continue est décrite puis mise en place. Cette dernière étape est décrite dans la sous-section suivante.

5.7.4 Amélioration continue

Juste avant le lancement de la plateforme, un plan de maintenance et d'amélioration continue est effectué afin de prévoir de futures améliorations. Ainsi, afin de permettre le développement en parallèle de nouvelles versions tout en assurant une bonne opération de l'outil, une copie de l'outil est effectuée. Cette copie est hébergée sur les mêmes serveurs que la version de production et sur tableau desktop et permet de prototyper sans impacter la version en ligne. Une fois les modifications testées ces dernières peuvent être implémentées. La figure 5.23 illustre cette architecture.

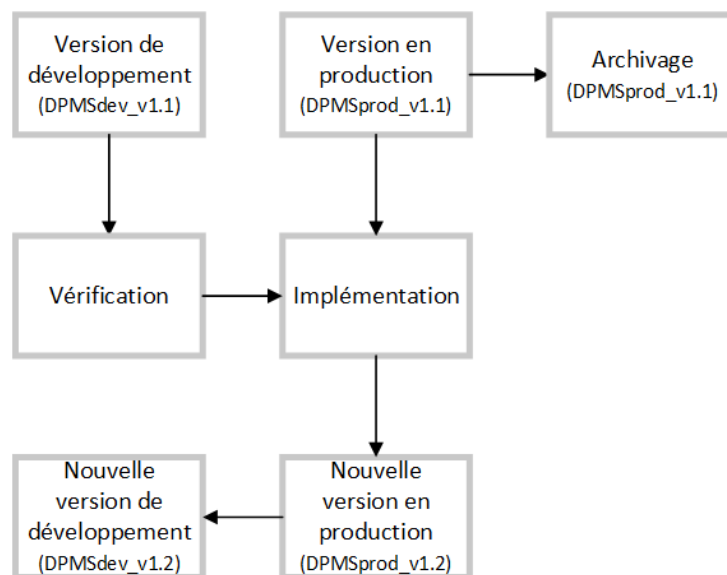


Figure 5.23 Gestion des versions de la plateforme

Ainsi, la gestion du développement de l'outil est réalisée en parallèle des opérations de la plateforme en ligne. Plus précisément, le processus décrit à la figure 5.23 est effectué sur

tableau pour l'outil de visualisation et sur Github pour le code source de la plateforme. L'architecture choisie en couche dans la section 5.3 permet d'ajouter de nouveaux modules à la plateforme tout en assurant la continuité des opérations. Une fois mise en production, de nouveaux attributs peuvent être ainsi ajoutés en moins de 24h sans nécessiter la réorganisation du code source de l'outil. Cette étape finalise alors le développement d'une plateforme collaborative visant à l'amélioration des performances de production des PME constituant un réseau décentralisé.

CHAPITRE 6 CONCLUSION ET RECOMMANDATIONS

6.1 Synthèse des travaux

Le but poursuivi dans le cadre de ce projet de recherche est le développement d'une méthodologie permettant de construire des outils collaboratifs visant à l'amélioration des performances de production d'un réseau décentralisé de PME.

L'état de l'art effectué à la section 2 a permis de mettre en évidence la diminution des coûts d'utilisation des technologies d'industrie 4.0.. Cette diminution permet la réalisation d'une transition de ces technologies vers les PME. En particulier, il a été montré que ces PME utilisent déjà des techniques de collecte de données de l'industrie 4.0.. De plus, il est démontré que les entreprises des réseaux de PME décentralisés possèdent de nombreux points communs en termes de gestion des performances tels que leurs caractéristiques géographiques et économiques.

Chaque entreprise souhaite améliorer ses performances de production en diminuant son isolement par rapport aux autres acteurs du réseau. Il existe, dès lors, la nécessité pour ces PME de créer des outils basés sur la valorisation de données. Si quelques outils d'optimisation de performances ont déjà été théorisés dans la littérature, il existe peu d'outils collaboratifs pour les PME. À partir des éléments de la revue de littérature et de l'identification des besoins des réseaux de PME, une méthodologie permettant de développer un outil collaboratif visant à l'amélioration des performances de production d'un réseau de PME a été établie.

Les étapes principales de ce processus méthodologique sont les suivantes. Dans un premier temps, une compréhension globale du contexte industriel est effectuée. Ensuite, un processus de collecte de données est organisé. Les données collectées sont préparées et nettoyées pour être présentées à l'aide d'un outil de visualisation. Par la suite, des indicateurs de comparaison et de performance sont générés afin d'améliorer les performances des entreprises du réseau. Enfin les outils construisent son déployé et validé suivant un processus décrit dans ce mémoire.

Afin de démontrer l'approche pratique de cette méthodologie ainsi que son succès, la méthodologie proposée est appliquée au cas d'étude de l'industrie agroalimentaire au Canada et plus précisément au secteur de l'industrie fourragère. Des outils de présentation et d'amélioration des performances de ces PME sont développés à l'aide de la méthodologie proposée puis implémentés.

Finalement, l'apport concret de ce travail est la méthodologie de développement d'un outil

d'aides à la décision collaboratif pour les réseaux décentralisés de PME. À la fin de ce travail de recherche, l'outil conçu à partir de la méthodologie proposé est développé et une première itération est déployée pour un réseau de ferme au Canada. Cet outil permet alors pour ces exploitants de connaître leurs performances réelles de production ainsi que d'obtenir des informations permettant d'améliorer leurs performances de production.

6.2 Limitations de la solution proposée

La conception et le déploiement de l'outil présenté dans ce mémoire à permis d'identifier certaines limitations. L'outil développé est constitué d'un processus de collecte de données semi-automatisé qui dépend fortement de collecte manuelle de données. Ce choix de conception est adapté dans le cadre d'un prototypage, mais peut entraîner de grandes répercussions sur les étapes de traitements de données. Si des étapes rigoureuses de préparation des données ont été effectuées, le processus actuel peut être amélioré et devrait être automatisé.

L'outil de comparaison développé est effectué suivant des critères géographiques uniquement et n'est pas réalisé suivant des groupements par des caractéristiques similaires. De plus, l'architecture de l'outil n'est pas pensée pour opérer en temps réel ce qui peut induire certaines limitations pour l'utilisateur final.

L'architecture du code de la plateforme est effectuée par une méthodologie de type couche ce qui a permis un développement rapide de l'outil, mais n'est pas très adapté pour les futurs besoins du projet. Enfin, les sources de données sont trop nombreuses et sous des formats et des structures différentes, ce qui complique l'architecture de la plateforme, et augmente la probabilité d'une déféctuosité. Une solution possible pour résoudre ce défi technologique serait donc de limiter le nombre de sources de données. Ainsi, la comparaison des performances est pour l'instant seulement effectuée en fonction d'un critère géographique.

6.3 Améliorations futures

L'ensemble des limitations énumérées dans la section précédente permet d'identifier quelques perspectives d'amélioration pour le futur développement de l'outil.

Les prochains objectifs de ce sujet de recherche viseront à l'optimisation globale du réseau d'entreprise. Ensuite, d'autres modules permettant de communiquer et d'influencer de bonnes pratiques aux réseaux d'entreprises seront développés. L'Analyse des entreprises les plus performantes, de leurs pratiques ainsi que des entreprises les moins performantes du réseau pourrait permettre d'optimiser les performances des PME. Ainsi, la comparaison des perfor-

mances de chaque entreprise avec les entreprises voisines pourrait être utilisée pour faciliter l'implantation de nouvelles mesures et limiter la résistance au changement des exploitants de ses PME.

Une autre perspective possible est l'identification des facteurs ayant le plus d'impact négatif sur les baisses de rendement. Le développement de ce module permettrait de quantifier l'impact de ces facteurs et d'améliorer les performances du réseau global de PME. Ensuite, le développement d'un nouveau système basé sur une architecture modulaire et en temps réel afin de permettre aux producteurs d'utiliser directement l'outil lorsqu'ils travaillent. Le développement des modules de prédiction et d'estimation des rendements de production obtenus pour l'ensemble du cycle d'exploitation sera par la suite étudié. Ces modules permettront de conseiller les PME afin d'améliorer leurs pratiques et leurs rendements de productions.

Enfin trois outils présentent des voies intéressantes dans le cadre de ce projet. Tout d'abord, un module d'analyses permettant aux gestionnaires et aux gouvernements d'influencer la stratégie globale du réseau à l'échelle nationale pourrait être conçu. Puis, un outil de suivi de l'évolution du réseau global permettant de monitorer la stratégie de production des entreprises peut être développé. Finalement, l'utilisation d'outils de segmentation afin de regrouper les agriculteurs suivant différentes similarités sera étudiée. L'ensemble de ces modules contribuera ainsi à la réalisation d'un outil d'aide à la décision et de gestion des performances d'un réseau décentralisé de PME.

RÉFÉRENCES

- [1] Charu C Aggarwal, Mansurul A Bhuiyan, and Mohammad Al Hasan. Frequent pattern mining algorithms: A survey. In *Frequent pattern mining*, pages 19–64. Springer, 2014.
- [2] Lara Agostini, Roberto Filippini, and Anna Nosella. Management and performance of strategic multipartner sme networks. *International Journal of Production Economics*, 169:376–390, 2015.
- [3] Lara Agostini, Roberto Filippini, and Anna Nosella. Management and performance of strategic multipartner sme networks. *International Journal of Production Economics*, 169:376–390, 2015.
- [4] V. Alcácer and V. Cruz-Machado. Scanning the industry 4.0: A literature review on technologies for manufacturing systems. *Engineering Science and Technology, an International Journal*, 22(3):899–919, 2019.
- [5] Vladimir Jovan Andrej Rakar, Sebastjan Zorzu. Assesment of production performance by means of kpi. *Control*, 2004.
- [6] Alvine B Belle, Ghizlane El Boussaidi, Timothy C Lethbridge, Segla Kpodjedo, Hafedh Mili, and Andres Paz. Systematically reviewing the layered architectural pattern principles and their use to reconstruct software architectures. *arXiv preprint arXiv:2112.01644*, 2021.
- [7] Pete Chapman, Julian Clinton, Randy Kerber, Thomas Khabaza, Thomas Reinartz, Colin Shearer, Rudiger Wirth, et al. Crisp-dm 1.0: Step-by-step data mining guide. *SPSS inc*, 9:13, 2000.
- [8] Nilubon Chonsawat and Apichat Sopadang. The development of the maturity model to evaluate the smart smes 4.0 readiness. In *Proceedings of the international conference on industrial engineering and operations management*, pages 354–363, 2019.
- [9] Aman Deep, Peter Guttridge, Samir Dani, and Neil Burns. Investigating factors affecting erp selection in made-to-order sme sector. *Journal of Manufacturing Technology Management*, 2008.
- [10] John J Dudley and Per Ola Kristensson. A review of user interface design for interactive machine learning. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 8(2):1–37, 2018.

- [11] Aldónio Ferreira and David Otley. The design and use of performance management systems: An extended framework for analysis. *Management Accounting Research*, 20(4):263–282, 2009.
- [12] Government of Canada. Key small business statistics — 2020. Technical report, 2020.
- [13] Hui Han and Silvana Trimi. Towards a data science platform for improving sme collaboration through industry 4.0 technologies. *Technological Forecasting and Social Change*, 174:121242, 2022.
- [14] Steffen Huber, Hajo Wiemer, Dorothea Schneider, and Steffen Ihlenfeldt. Dmme: Data mining methodology for engineering applications – a holistic extension to the crispdm model. *Procedia CIRP*, 79:403–408, 2019. 12th CIRP Conference on Intelligent Computation in Manufacturing Engineering, 18-20 July 2018, Gulf of Naples, Italy.
- [15] Jacob Høxbroe Jeppesen, Emad Ebeid, Rune Hylsberg Jacobsen, and Thomas Skjødeberg Toftegaard. Open geospatial infrastructure for data management and analytics in interdisciplinary research. *Computers and Electronics in Agriculture*, 145:130–141, 2018.
- [16] P. Ravi Kumar, P. Herbert Raj, and P. Jelciana. Exploring data security issues and solutions in cloud computing. *Procedia Computer Science*, 125:691–697, 2018. The 6th International Conference on Smart Computing and Communications.
- [17] Trudie Lang. Advancing global health research through digital technology and sharing data. *Science*, 331(6018):714–717, 2011.
- [18] In Lee. Big data: Dimensions, evolution, impacts, and challenges. *Business Horizons*, 60(3):293–303, 2017.
- [19] Zixu Liu, Pedro Sampaio, Grigory Pishchulov, Nikolay Mehandjiev, Sonia Cisneros-Cabrera, Arnd Schirrmann, Filip Jiru, and Nisrine Bnouhanna. The architectural design and implementation of a digital platform for industry 4.0 sme collaboration. *Computers in Industry*, 138:103623, 2022.
- [20] Abdul Majeed and Sungchang Lee. Anonymization techniques for privacy preserving data publishing: A comprehensive survey. *IEEE Access*, 9:8512–8545, 2020.
- [21] Sameer Mittal, Muztoba Ahmad Khan, David Romero, and Thorsten Wuest. A critical review of smart manufacturing & industry 4.0 maturity models: Implications for small and medium-sized enterprises (smes). *Journal of Manufacturing Systems*, 49:194–214, 2018.

- [22] Kevin Nagorny, Pedro Lima-Monteiro, Jose Barata, and Armando Walter Colombo. Big data analysis in smart manufacturing: A review. *International Journal of Communications, Network and System Sciences*, 10(3):31–58, 2017.
- [23] Ngakan Nyoman Kutha Krisnawijaya, Bedir Tekinerdogan, Cagatay Catal, and Rik van der Tol. Data analytics platforms for agricultural systems: A systematic literature review. *Computers and Electronics in Agriculture*, 195:106813, 2022.
- [24] OECD. Enhancing the contributions of smes in a global and digitalised economy. *Paris. Retrieved Feb, 25:2019*, 2017.
- [25] Konstantinos Perakis, Fenareti Lampathaki, Konstantinos Nikas, Yiannis Georgiou, Oskar Marko, and Jarissa Maselyne. Cybele – fostering precision agriculture & livestock farming through secure access to large-scale hpc enabled virtual industrial experimentation environments fostering scalable big data analytics. *Computer Networks*, 168:107035, 2020.
- [26] Dorian Pyle. *Data preparation for data mining*. Morgan Kaufmann, 1999.
- [27] Lei Ren, Lin Zhang, Fei Tao, Chun Zhao, Xudong Chai, and Xinpei Zhao. Cloud manufacturing: from concept to practice. *Enterprise Information Systems*, 9(2):186–209, 2015.
- [28] Nitin Sawant and Himanshu Shah. *Big Data Application Architecture*, pages 9–28. Apress, Berkeley, CA, 2013.
- [29] Doruk Sen, Melike Ozturk, and Ozalp Vayvay. An overview of big data for growth in smes. *Procedia - Social and Behavioral Sciences*, 235:159–167, 2016. 12th International Strategic Management Conference, ISMC 2016, 28-30 October 2016, Antalya, Turkey.
- [30] Jan Stentoft, Kent Aadsbøll Wickstrøm, Kristian Philipsen, and Anders Haug. Drivers and barriers for industry 4.0 readiness and practice: empirical evidence from small and medium-sized manufacturers. *Production Planning & Control*, 32(10):811–828, 2021.
- [31] Jeremy Straub. A modern blackboard architecture implementation with external command execution capability. *Software Impacts*, 11:100183, 2022.
- [32] Dennis Trotta and Patrizia Garengo. Assessing industry 4.0 maturity: An essential scale for smes. In *2019 8th International Conference on Industrial Technology and Management (ICITM)*, pages 69–74, 2019.

- [33] Nikos Tsikriktsis. A review of techniques for treating missing data in om survey research. *Journal of operations management*, 24(1):53–62, 2005.
- [34] Dan Undersander and Dennis Cosgrove. *Alfalfa management guide*. American Society of Agronomy Crop Science Society of America Soil Science . . . , 2011.
- [35] B. Verhaelen, F. Mayer, S. Peukert, and G. Lanza. A comprehensive kpi network for the performance measurement and management in global production networks. 15, 2021.
- [36] Agostino Villa. *Managing cooperation in supply network structures and small or medium-sized enterprises*. Springer, 2014.
- [37] Tom Wilcox, Nanlin Jin, Peter Flach, and Joshua Thumim. A big data platform for smart meter data analytics. *Computers in Industry*, 105:250–259, 2019.
- [38] Ilaria Zambon, Massimo Cecchini, Gianluca Egidi, Maria Grazia Saporito, and Andrea Colantoni. Revolution 4.0: Industry vs. agriculture in a future development for smes. *Processes*, 7(1), 2019.

ANNEXE A MÉTHODOLOGIE COMPLÈTE

Cette annexe A présente la méthodologie complète décrite dans le cadre de ce projet de recherche. Toutes les grandes étapes y sont détaillées.

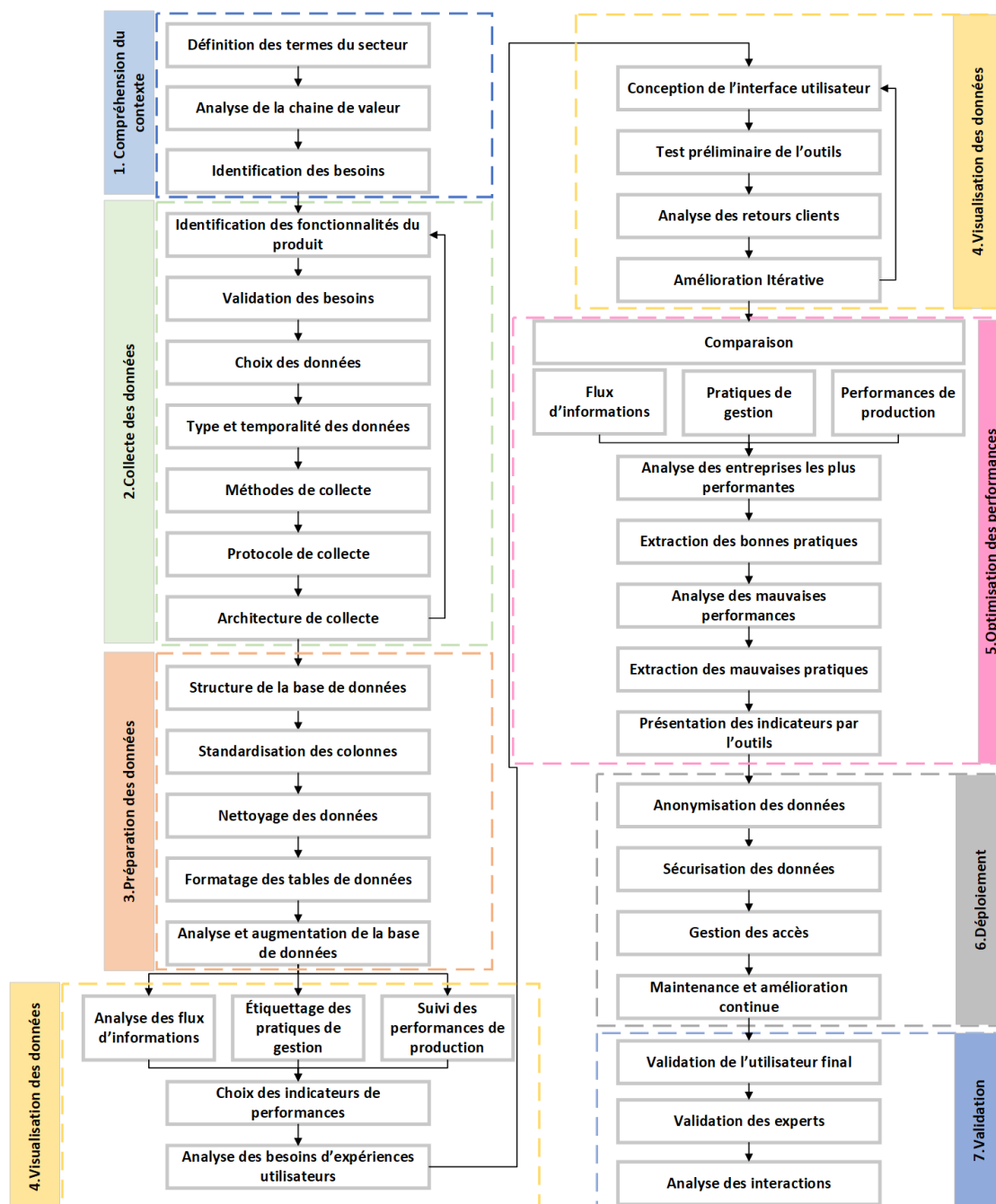


Figure A.1 Méthodologie complète de développement de l'outil collaboratif