

Élaboration d'un outil d'aide à la décision basé sur les règles d'association pour une PME

LAURE SIRET^{1,2}, CHRISTOPHE DANJOU², BRUNO AGARD^{1,2}

¹ Laboratoire en Intelligence des Données
Département de mathématiques et génie industriel,
École Polytechnique de Montréal, CP 6079, succursale Centre-Ville, Montréal, Québec, Canada
laure.siret@polymtl.ca, christophe.danjou@polymtl.ca, bruno.agard@polymtl.ca

² Centre Interuniversitaire de Recherche sur les Réseaux d'Entreprise, la Logistique et le Transport (CIRRELT)

Résumé – L'amélioration des moyens de communication sur lesquels repose l'industrie 4.0 entraîne une collecte de données de production massive en industrie. À ce titre, l'utilisation du data mining se démocratise et encourage le développement d'une production intelligente avec un contrôle en temps réel. Cependant, les informations capturées au cours de la production doivent être stockées et triées avant l'utilisation de n'importe quelle fonction de data mining. La problématique pour les Petites et Moyennes Entreprises (PME) est donc de savoir comment traiter leurs données pour améliorer la prise de décisions pour la production. À cette fin, nous proposons une méthodologie en quatre étapes qui permet de structurer les données importantes, d'identifier et d'utiliser des indicateurs de performance (KPI) et finalement, d'analyser les données pour générer des rapports sur la production.

Abstract – The improvement of the information and communication technologies on which Industry 4.0 relies leads to a mass production data collection in industry. As such, the use of data mining is democratized and encourages the development of intelligent production with real-time control. However, information captured during production must be stored and sorted before any use of any data mining function. Thus, how to process data to improve the production decision making is a major problem for small and medium-sized enterprises (SME). Therefore, we propose a four-step methodology that allows structuring important data, identifying and using Key Performance Indicators (KPIs) and finally analyzing data and generating reports on production.

Mots clés – Industrie 4.0, KPI (Indicateur de Performance), OAD (Outil d'Aide à la Décision), Règles d'associations, Systèmes d'information.

Keywords – Industry 4.0, KPI (Key Performance Indicator), DST (Decision Support Tool), Association rules, Information system.

1 INTRODUCTION

Depuis plus de 10 ans, la recherche s'intéresse à la fouille de données au sein des entreprises [Choudhary et al., 2009]. Packianathera et al. (2017) montrent que les enjeux économiques y sont très importants. La valorisation des données peut y prendre de nombreuses formes. Elle passe de l'analyse de profils clients [Packianathera et al., 2017] à l'étude de la qualité des produits et procédés [Köksal et al., 2011]. Ce dernier ilot se décompose en plusieurs grands secteurs, par exemple : la description du produit, sa classification, la prédiction de défauts et l'élaboration de paramètres d'optimisation [Köksal et al., 2011].

À cette multiplicité de méthodes déjà existantes, on ajoute le contexte actuel de l'industrie 4.0. Ce contexte porté par les concepts de la transparence de l'information et l'étude de l'environnement de production démocratise l'utilisation de la donnée [Kagermann 2015]. Il y a ainsi une explosion de la quantité, mais aussi du type de données qui peuvent être récupérées [Danjou et al., 2017]. On parle maintenant de « big data » et l'information ne concerne plus uniquement les données de production, mais l'environnement complet de l'usine [Witten et al., 2017].

Le besoin d'étudier ces données non interprétables à l'œil humain devient d'autant plus important et représente un réel enjeu stratégique pour les entreprises [Rissoan et Jouin, 2018]. Cependant, l'augmentation de la quantité de données, ordonnées parfois de manière nébuleuse [Lemberger et al., 2016], complexifie le choix d'une méthode adéquate pour les entreprises et là où le bon sens et l'expérience étaient autrefois suffisants, il faut désormais un analyste. Cela représente un défi pour les entreprises avec peu de moyens.

Un enjeu actuel pour les PME est l'amélioration de l'exploitation de leurs outils de production. Pour cela, une meilleure connaissance et un meilleur suivi de ce qui se passe sur le plancher de production est alors nécessaire. Ceci se traduit par un volume croissant de données à traiter. L'exploitation de celles-ci représente un challenge pour les PME qui n'ont pas forcément d'analystes. S'agissant de nouveaux métiers et techniques, le coût associé à ces analyses est difficile à supporter. Généralement, les PME ne possèdent pas encore les ressources, les outils et les méthodes pour répondre aux enjeux de l'exploitation des données. Dans le présent article, nous proposons une méthodologie de création d'un outil d'aide à la décision (OAD) ayant pour objectif

l'amélioration de l'exploitation des outils de production pour les PME, à partir de l'analyse des historiques de données. Cette méthodologie débute avec l'acquisition et le formatage des données d'un système ERP (Entreprise Ressources Planning) et va jusqu'à l'émission d'un rapport liant les paramètres de production et les aléas de production. La méthodologie s'appuie sur les outils d'analyses de bases de données (BDD), sur la mise en place d'indicateurs de performance et enfin sur l'intelligence artificielle via l'utilisation des règles d'association. Le cas spécifique d'une entreprise du secteur mécanique possédant de grandes quantités des données est analysé.

La structure de l'article est la suivante : la section 2 présentera un état de l'art qui montrera la place des données dans l'industrie 4.0, certains des outils de data mining puis leur lien avec l'industrie 4.0. La section 3 exposera la méthodologie proposée en 4 étapes. La section 4 exposera le cas d'étude qui montrera la mise en œuvre de la méthodologie. Cette étude de cas a été réalisée dans une entreprise du secteur mécanique et aéronautique. Finalement, la section 5 rappellera les résultats et limites de la méthode utilisée et proposera quelques perspectives.

2 ÉTAT DE L'ART

2.1 Industrie 4.0

2.1.1 Définitions

Tout comprendre de son environnement, avoir des produits modifiables et personnalisables rapidement selon la demande du client, tel est le contexte dans lequel apparaissent les premiers signes d'une 4^e révolution industrielle aussi présentée comme Industrie 4.0 en Allemagne dès 2011 [Drath et Horch, 2014].

Selon Halaška et Šperka (2018), une des caractéristiques de l'industrie 4.0 est l'évolution d'unité isolée et automatisée en unité connectée à son environnement de travail, automatisée et optimisant ses actions. Pour arriver à cela, Schuh et al. (2015) proposent six points clés de cette révolution : (1) la disponibilité en temps réel des informations, (2) l'utilisation de « smart products », (3) la mise à jour perpétuelle de la représentation des produits, (4) le prétraitement automatique des données de production, (5) l'intégration d'informations sur l'environnement extérieur et (6) le développement d'interfaces hommes-machines. Le « smart product » est un produit muni de la technologie AIDC (Automatic Identification and Data Capture) c'est à dire de capteurs localisables en tout temps et stockant les informations relatives à chaque étape de la production du produit [Schuh et al., 2015].

La donnée est donc placée au centre du concept de l'industrie 4.0. Les technologies de Cyber Physical System (CPS), Internet Of Things (IoT), People (IoP) et Services (IoS), sur lesquelles se base l'Industrie 4.0 servent à l'acquisition, au transfert, au stockage ou au traitement des informations.

2.1.2 Les capteurs et données

L'Industrie 4.0 implique donc, avant tout, une phase de surveillance et d'acquisition de données. Danjou et al. (2017) y réfèrent comme une base nécessaire avant toute progression vers d'autres niveaux du 4.0. Selon Nemeth et Peterkova (2018), beaucoup de ces données sont purement stockées et n'ont qu'un but d'historique non exploité. Dobos et al. (2018) présentent alors des applications possibles pour ce « Big Data ». Les méthodes de data mining sont particulièrement bien adaptées pour ce type d'études. Dans le contexte de l'Industrie 4.0, la notion de contrôle

en temps réel est primordiale or la gestion de ce contrôle peut être facilitée par la visualisation d'informations synthétiques ou d'indicateurs.

2.1.3 La recherche d'indicateurs

Les indicateurs de performances s'utilisent principalement dans les entreprises dès que l'on cherche à contrôler la productivité d'un environnement [Parmenter, 2015]. Un KPI (Key Performance Indicators) permet de quantifier une entité, un processus de production. Il existe plusieurs méthodes de génération de KPI et de nombreux articles se consacrent sur leur génération, chacun y aborde une thématique différente : l'uniformisation des connaissances d'une structure de production [Grabot, 1998], le résultat d'une analyse systémique [Tamim et al., 2018] ou encore l'impact de la combinaison et l'agrégation d'indicateurs [Hély et Antoni, 2019] en sont quelques exemples. Pour la suite de notre étude, la décision de suivre les premières étapes présentées par Grabot, (1998) et les conseils généraux présentés par Parmenter (2015) est prise. Les premières étapes de la conception d'une structure hiérarchique, présentée par Grabot, (1998), permettent de définir qu'un indicateur local (IL) permet de contrôler un objectif local, mesurant un processus. Un indicateur global (IG) surveille un objectif global et mesure une ressource. Un objectif global est le résultat d'une combinaison d'objectifs locaux. Parmenter (2015), permet de définir qu'un indicateur local est adimensionné, simple à interpréter et s'inscrit dans le temps.

Les KPI sont une part essentielle des données utilisées pour le data mining. Selon l'étude de la littérature, la majorité des articles utilisant le data mining dans l'industrie pour de l'analyse de rentabilité ou de qualité de produit industriel ne précisent pas comment les données d'étude ont été sélectionnées. Celles-ci sont souvent fournies au data analyste (personne responsable de l'analyse de données) par le data architecte (personne responsable de la structure des données) [Rissoan et Jouin, 2018]. La structuration et l'entreposage des données sont donc un point essentiel du 4.0.

2.1.4 Le data warehouse

Les informations mesurées par les différents capteurs sont stockées et organisées, dans ce qui s'appelle souvent un « data warehouse » ou « entrepôt » de données. Ces entrepôts se composent de plusieurs bases de données. Kimball et Ross (2013) définissent qu'une base de données se compose de deux grands types de tables. Le premier, ou « fact table », contient des individus ou listes d'événements et a comme entrées (ou colonnes) un certain nombre de paramètres codifiés. Le second, ou « dimension table », représente les tables contenant les références de ces codifications. La séparation des informations en plusieurs tables a pour but un meilleur partage des ressources entre les différents acteurs de l'entreprise ainsi qu'un accès plus rapide selon le type d'information. Un entreposage efficace des données facilite leur traitement, via notamment des fonctions du data mining.

2.2 Data Mining

Comme exposé dans la partie précédente, l'usine 4.0 génère beaucoup de données. Celles-ci ne sont pas facilement interprétables et se rapproche davantage du Big Data que du Smart Data. Là où le Smart Data concerne des regroupements de peu de données facilement visualisables et interprétables, le Big Data

nécessite l'utilisation d'outils d'analyses, en particulier des outils tels que les fonctions de data mining [Rissoan et Jouin, 2018].

2.2.1 Définition

Witten et al. (2017) définissent le Data Mining (DM) comme « le processus de découverte de patrons au sein de données. ». Le DM cherche à expliquer des relations, entre plusieurs données, qui ne sont pas visibles instantanément. Ces découvertes reposent sur une multitude de fonctions et algorithmes mathématiques présentés dans la partie suivante et doivent s'intégrer dans le contexte actuel du 4.0.

2.2.2 Les grandes familles d'outils

Il est possible de classer les fonctions de DM selon deux grandes familles qui sont les fonctions descriptives et les fonctions prédictives [Köksal et al., 2011] :

- La famille des fonctions descriptives se compose des fonctions de segmentation, d'association, d'agrégation et la recherche de patrons fréquents. Elles ont pour but la mise en relief de règles, relations ou caractéristiques dans les données. Ces résultats peuvent servir à la déduction de paramètres de production critiques [Cheng et al., 2018].
- La famille des fonctions prédictives comprend, la classification, la prédiction et le traitement de séries temporelles. Elles ont pour principal objectif la recherche de tendances pertinentes dans les données. Les fonctions prédictives s'utilisent afin d'anticiper l'avenir et ainsi tenter de fournir un avantage concurrentiel aux entreprises [Cheng et al., 2018].

On note que les recherches portant sur l'utilisation du DM pour l'industrie de l'usinage s'orientent majoritairement sur de la prédiction [Köksal et al., 2011]. Dans le cas où l'entreprise souhaite avoir un retour sur sa production, les fonctions descriptives sont les plus adéquates [Choudhary et al., 2009].

Parmi les différents outils disponibles, pour la suite de l'étude, seules les **règles d'association** seront utilisées. En effet, l'objectif n'est pas de regrouper des produits ou processus systématiquement problématiques, le « clustering » et le « sequence pattern mining » [Cheng et al., 2018] ne sont donc pas pris en compte. Les fonctions de règles d'association sont sélectionnées. Les règles d'association ne sont pas très présentes lors d'études d'analyse de qualité mais sont souvent utilisées pour des analyses de marché et de consommation client (le problème du « market basket » [Linoff et Berry, 2011]) ou pour le design de produit [Choudhary et al., 2009]. Il est alors intéressant d'observer leur utilisation dans le contexte de la production industrielle.

On rappelle ici le principe des règles d'association. Une règle est un ensemble de deux membres. Le premier (M_1) et le second (M_2) sont composés d'un ou plusieurs éléments appelés items, appartenant à une même table de données et $M_1 \neq M_2$. Une règle d'association traduit la cooccurrence de M_1 et M_2 et se traduit formellement par la relation $M_1 \rightarrow M_2$.

Une règle est caractérisée par 3 paramètres : la confiance, le support minimal et maximal et le « lift ». Ils concernent respectivement, la certitude de la règle, la proportion d'individus touchés au minimum ou au maximum et la corrélation entre M_1 et M_2 [Buddhakulsomsiri et al., 2006].

La méthode la plus utilisée pour la génération de règles d'associations est la méthode « APriori » [Linoff et Berry, 2011].

2.2.3 Données et DM en industrie 4.0

Les études par des fonctions de DM dans le milieu industriel reposent principalement sur deux types de données pouvant être collectées : les données mesurant une réalité physique (vitesse, effort, débit) ou les données système (message de connexion, déconnexion). Nemeth et Peterkova (2018) indiquent que dans le cadre de l'industrie 4.0 le second type a tendance à être favorisé au premier car, les entreprises se préoccupent davantage de la partie technologique et usine autonome de l'Industrie 4.0. À ce titre, les entreprises s'équipent de technologies qui captent automatiquement des données logicielles, mais ces mêmes entreprises ne cherchent pas nécessairement à récupérer les informations présentes dans les automates programmables des machines implantées. Les points clés (2) et (5) de Schuh et al. (2015) qui sont l'utilisation de « smart product » et l'intégration d'informations sur l'environnement extérieur, se reposent essentiellement sur l'acquisition des données physiques et ne sont pas satisfaits dans ce cas. Afin de remédier à la situation où seules les données système seraient étudiées, Nemeth et Peterkova (2018) et Vitalii et al. (2018) proposent des méthodologies d'acquisition de données pour respectivement l'industrie automobile et métallurgique.

L'étude de la littérature fournit ainsi plusieurs parties applicables pour la mise en place d'un OAD mais pas de méthode générique permettant l'étude de données de production depuis leur collecte jusqu'à la génération de rapports de production.

3 METHODOLOGIE DE DEVELOPPEMENT D'UN OUTIL D'AIDE A LA DECISION DANS UNE PME

Nous proposons une méthodologie de développement d'un outil d'aide à la décision dans le contexte industriel d'une PME. Cet outil repose sur l'analyse de données de production afin d'y découvrir des paramètres de production critiques.

3.1 Définition des termes

On appelle **base de données** un ensemble de tables de données. Chaque **table de données**, notée T_i , se compose de plusieurs colonnes et de plusieurs lignes. La nature des données stockées peut-être : une chaîne de caractères, un entier, une date ou un booléen soit respectivement de nature SQL VARCHAR, INTEGER, DATETIME ou DATE et BOOLEAN.

La **cartographie des tables** sera appelée diagramme relationnel. Une **clé de recherche**, notée K_i , est une colonne commune à plusieurs tables de la base de données. Cette colonne commune n'a pas nécessairement le même en-tête d'une table à l'autre. Les K_i sont identifiables via le diagramme relationnel.

On appelle **chemin** une liste de plusieurs clés K_i permettant le passage d'une table T_i à une table T_k .

Un chemin est dit actif lorsqu'il est usuellement utilisé par la société pour l'accomplissement de rapport ou autres requêtes de travail.

Un **rapport** est un « output » ou message informatique généré par le système et faisant suite à une requête utilisateur ou à une action du système (mise à jour, maintenance).

Nous appelons **table des indicateurs** ($T_{\text{indicateurs}}$), une table que l'on crée et contenant des informations extraites de T_i de la base de données.

3.2 Méthodologie proposée

La méthodologie se compose de quatre phases qui seront détaillées par la suite. L'exécution complète de la méthodologie

permet une meilleure compréhension de l'architecture de la base de données ainsi que la création d'une cartographie efficace des différentes tables, la mise en place d'indicateurs, leur lecture et interprétation au quotidien et finalement la mise en place d'un outil d'aide à la décision reposant sur des règles d'associations. Les différentes étapes de la méthodologie ainsi que l'ensemble des documents utilisés ou créés sont synthétisés Figure 1. Les documents utilisés durant l'étude (rapports entreprise, diagrammes relationnels, listes de contraintes, etc.) peuvent être utilisés en lecture seule, écriture seule ou lecture et écriture.

3.3 Phase I : Cartographie des données

Cette première phase a pour but d'organiser un ensemble de tables d'une même base de données. Dans le cas où les tables sont de tailles différentes, et n'ont pas les mêmes fonctions au sein de l'entreprise, il est important de pouvoir les trier relativement les unes aux autres, mais également de manière absolue au sein de la base.

L'ensemble de ces tables sont liées entre elles par une ou plusieurs clés. Une table peut être liée à une ou plusieurs autres tables. Idéalement, il n'y aurait qu'un seul chemin pour passer de T_i à T_k mais dans la réalité, il y a souvent, plusieurs chemins. De plus, tous les chemins ne sont pas nécessairement actifs. Cette problématique vient de la maintenance de la base de données (BDD) dans son ensemble.

Étape I.1 : Réduction de l'espace de travail

Dans un premier temps, nous cherchons à réduire l'espace de travail pour ne garder que les bases essentielles au fonctionnement de l'entreprise. L'identification des bases essentielles passe par l'analyse de rapports de contrôle générés par l'entreprise. La réduction ne contient que les bases présentes lors de l'exécution de n'importe quelle requête sur un produit ou machine.

Étape I.2 : Sectorisation des tables

Ces tables essentielles sont ensuite réorganisées par domaine. Chaque regroupement de tables est appelé « ilot ». Ces ilots sont fixés arbitrairement et dépendent du contexte industriel. Les produits, la gestion des non-conformités ou des processus restent des domaines communs à un grand nombre d'industries. Par exemple, le domaine des « produits » s'organisera avec *a minima* l'ensemble des tables nécessaires à l'obtention d'une clé produit unique, du numéro de lot et de la gamme standard.

Étape I.3 : Cheminement dans la cartographie

On rappelle qu'un chemin est une liaison entre deux tables de BDD. Dans les BDDs de production, il est fréquent que certains d'entre eux soient inactifs et que cette inactivité ne soit pas visible dans le diagramme relationnel. Pour une optimisation du temps de recherche et pour une cohérence de la cardinalité des données obtenues, il est nécessaire de toujours passer par les mêmes tables intermédiaires. Le diagramme relationnel est mis à jour par la suppression des liens qui ne sont plus actifs ou qui retournent des cardinalités absurdes. Une cardinalité absurde équivaut à un nombre d'individus différent pour une requête censée être équivalente pour la base de données et peut s'expliquer par une mauvaise codification de la clé ou une transformation des données saisies au cours du temps sans qu'il y ait eu modification de la clé. Enfin, un seul chemin est choisi parmi l'ensemble des chemins actifs. Ce choix repose sur les retours des experts métier et l'analyse de fréquences d'utilisation des chemins.

Étape I.4 : Hypothèses et contraintes d'ilot

Les tables les plus fréquemment utilisées sont identifiées et sont classées relativement par ilot et de manière absolue et il existe un cheminement unique pour aller d'une table T_i à une table T_k .

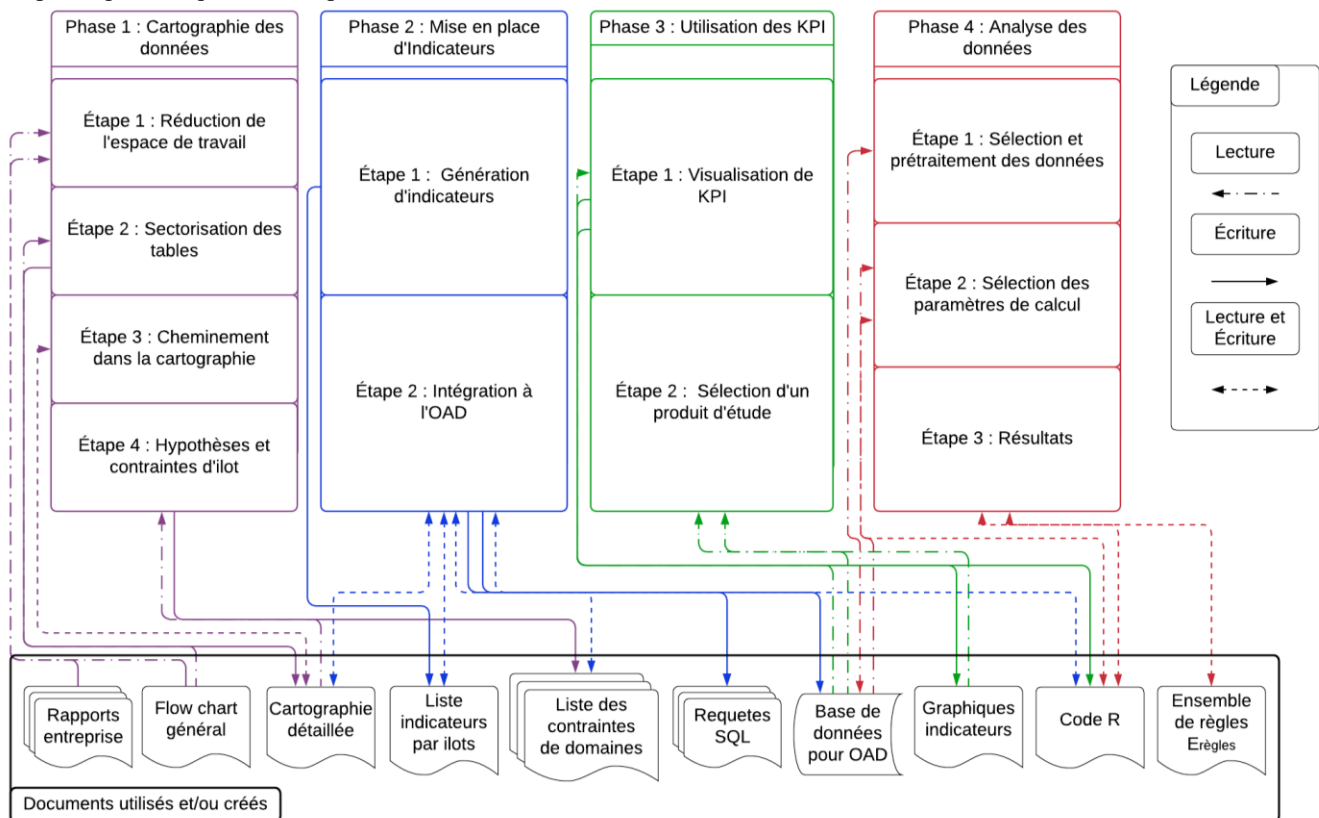


Figure 1 : Méthodologie d'implantation : Plan général

Chacune de ces tables contient des éléments non pertinents pour la suite de l'étude. Effectivement, certains individus peuvent être caducs en fonction du temps, être des tests de la production ou des tests du système informatique. Il est donc nécessaire de mettre en place une liste de contraintes pour chacune des tables. De manière analogue à l'étape 3.2.3 on peut regrouper des contraintes selon les îlots et avoir une liste de contraintes générale.

Exemple : la contrainte sur l'âge des données utilisables reposera sur la date de la dernière table créée. Le seul moyen d'augmenter l'historique serait de changer de chemin, entraînant alors une perte d'informations. Un équilibre entre la plus jeune des tables que l'on sélectionne dans la cartographie et l'historique perdu doit être trouvé. Les contraintes peuvent également venir de l'entreprise qui peut décider de ne pas faire d'apprentissage à partir de données sensibles ou qu'elle juge non fiables.

3.4 Phase II : Mise en place d'indicateurs

Rappelons ici que le but de la méthodologie est de permettre la conception d'un outil d'aide à la décision. Les indicateurs sont amenés à évoluer de manière constante en fonction de leur pertinence à l'environnement [Grabot, 1998]. Pour une pérennité de l'OAD, il faut pouvoir générer les indicateurs de manière homogène et les intégrer à l'OAD à n'importe quel instant.

Étape II.1 : Génération d'indicateurs

Pour que l'entreprise puisse modifier par elle-même les indicateurs, nous proposons de créer des indicateurs selon la méthodologie suivante :

- **Identifier** un domaine en besoin (îlot de la cartographie)
- **Déduire** une liste d'**indicateurs globaux** (IG)
- Pour chaque indicateur global, **déduire** une liste d'**indicateurs locaux** (IL)
- Pour chaque indicateur local, **identifier** l'ensemble idéal des **données brutes** et la **valeur interprétée** induite.
- **Identifier** l'ensemble disponible de données brutes et valeur interprétée induite et les ajouter à **T_{indicateurs}**.

Les trois premières étapes permettent d'orienter la recherche d'indicateurs du plus général au plus élémentaire, ce que [Grabot, 1998] définit comme une recherche hiérarchique. Les précédentes étapes permettent également de couvrir les indicateurs « globaux » (concernant les processus) et « locaux » (concernant les ressources). En effet, là où le processus impacte tout le plancher de production [Yahyaoui, 2015], les ressources concernent la performance d'une opération spécifique [Grabot, 1998]. Il est nécessaire d'avoir un bon équilibre de ces deux types d'indicateurs pour avoir une représentation réaliste du centre de production. Finalement, cette structure est utilisée pour l'analyse d'une performance passée. L'entreprise est libre de changer les indicateurs de plus bas niveau en fonction de l'information présente dans le système ce qui lui permet de reporter la de collecte de certaines données difficiles d'accès (étape 4) par d'autres en attendant une mise à jour du système d'information.

Étape II.2 : Intégration des informations à l'OAD

L'information idéalement à extraire des BDDs est identifiée. Il faut ensuite mettre à jour la cartographie et filtrer les données qui seront insérées à **T_{indicateurs}**. L'intégration à la cartographie est présentée sur la Figure 2. Ce schéma détaille en particulier le dernier point de la méthodologie de création d'indicateurs.

Chaque indicateur implique un prétraitement des tables. Le prétraitement passe par une distinction de différents cas d'études

détailés dans la Figure 3. Une première disjonction de cas est réalisée selon deux types de tables : celles contenant un champ avec des données temporelles et celles sans. Les **T_i** avec des données temporelles auront une itération de traitement supplémentaire afin de respecter le problème de cardinalité précédemment énoncé.

Dans le cas de la présence d'une donnée temporelle, cas E.2.A de la Figure 3, il existe deux cas d'étude :

- Cas E.2.A.1, si la donnée est postérieure à la date de création de la plus récente des tables, alors il faut mettre à jour la contrainte générale sur la date limite. Cela implique une réduction d'historique.
- Cas E.2.A.2 si la donnée est antérieure à la date limite générale, aucun traitement n'est effectué, la contrainte temporelle sur la table est la même que la contrainte générale.

Dans le cas E.2.B où il n'y a pas de données temporelles dans la table, nous passons à l'étude des données à extraire. Une liste de contraintes est créée au nom de la table. Pour la suite, il existe deux cas d'étude :

- Cas E.2.B.1 les données étudiées appartiennent à un ensemble de valeurs fini.
- Cas E.2.B.2 les données peuvent avoir un nombre infini de valeurs. « Un nombre infini de valeurs » est à prendre, ici, au sens d'une série arithmétique. Cela signifie que les données représentent un numéro de série, numéro de lot ou toute combinaison alphanumérique arbitraire, définie par l'entreprise pour suivre un produit, événement, etc.

Pour un ensemble fini de termes, la mise en place de contraintes est réalisée par : l'affichage de l'ensemble des éléments de la liste, l'identification des éléments à conserver ou à éliminer et la mise à jour de la liste des contraintes. Cette étape de visualisation est répétée jusqu'à ce qu'il ne reste que les individus pertinents.

Pour un ensemble infini, seuls les individus identifiés comme non réels ou aberrants sont éliminés. Les individus non réels peuvent être des essais de production ou de système. Les individus aberrants sont souvent identifiés lors de la correction des hypothèses par des experts de l'entreprise ou lors de la visualisation des premiers résultats. Le protocole de sélection des données est représenté Figure 3.

3.1 Phase III : Utilisation des KPI

La méthodologie proposée s'inscrit dans un contexte manufacturier. Une entreprise a des objectifs stratégiques, il existe donc des aspects critiques de la production et certains indicateurs sont plus « importants » que d'autres. Ces indicateurs sont les KPIs et sont notés comme tels par la suite. Il est également probable qu'il existe déjà, avant la mise en place de l'OAD, des moyens de visualisation de résultats au sein de l'entreprise. L'OAD devrait alors s'intégrer comme un module complémentaire aux visualisations déjà présentes. À cette fin, il faut que les KPIs se représentent de la manière la plus compacte possible et avec le maximum d'informations.

Étape III.1 : Visualisation des KPI

Nous proposons une visualisation partielle de l'ensemble des indicateurs obtenus. Dans un premier temps, l'affichage d'un seul indicateur par îlot est suggéré. Cette limitation a pour justification la facilité de lecture par un ensemble de personnes qui peuvent être dirigeants ou opérateurs.

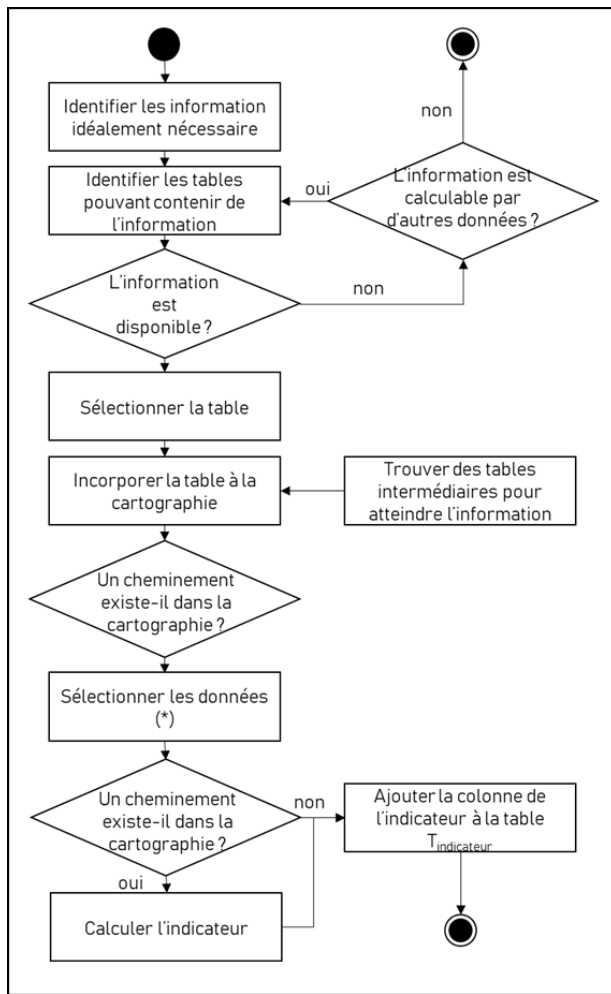


Figure 2. Intégration d'un indicateur à l'OAD

Dans un second temps, l'objectif d'une visualisation d'indicateurs est la comparaison avec sa référence. Cette référence peut se baser sur les objectifs stratégiques de l'entreprise via une valeur fixe et non directement corrélée au plancher de production ou elle peut représenter l'historique de production et se baser sur des valeurs passées de l'entreprise.

Nous rappelons ici que pour pouvoir calculer des résultats homogènes, l'ensemble des indicateurs sont adimensionnés.

Enfin, la visualisation peut être réalisée par un graphique contenant les valeurs de l'historique ou une icône bilan qui indiquera simplement l'état de l'indicateur vis-à-vis du seuil défini dans le paragraphe ci-dessus.

Étape III.2 : Sélection d'un produit d'étude

L'objectif des indicateurs de cet OAD est de pouvoir repérer des produits ou lots de produits problématiques.

Un produit est identifié comme intéressant à étudier lorsque l'un de ses indicateurs dépasse son seuil de tolérance. Sous cette condition, les produits suivront donc la phase 4 de la méthodologie.

3.1 Phase IV : Analyse des données

Pour cette partie, nous supposons avoir sélectionné à la phase 3 un produit ou une machine cible sur lequel réaliser une analyse de données. Cet objet d'étude est appelé : « Cible », noté C.

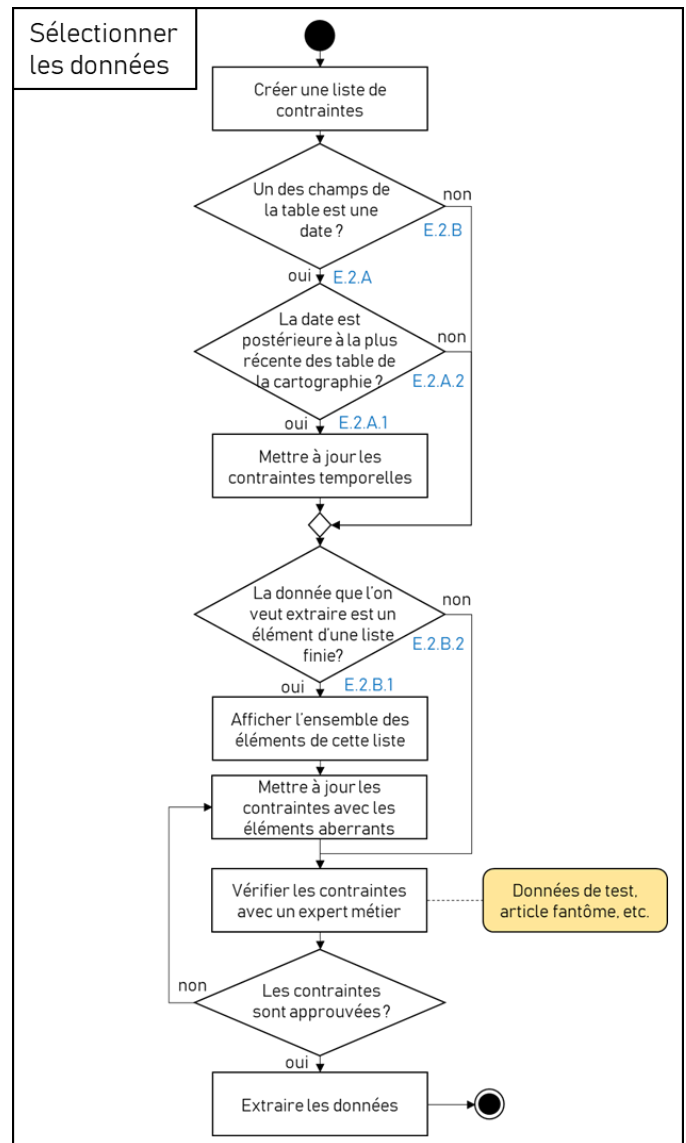


Figure 3. Sélection des données pour un indicateur

L'ensemble des données collectées sont rassemblées dans la table $T_{indicateurs}$. Le cas étudié est celui d'un OAD qui ne s'utilise qu'à des fins explicatives, autrement dit, l'OAD présenté, mettra en évidence des dysfonctionnements qui ont déjà eu lieu dans la production et se propose d'en donner une explication.

Étape IV.1 : Sélection et prétraitement des données

Pour la suite de notre étude, l'objectif n'est pas de regrouper des produits ou processus systématiquement problématiques ni d'analyser un processus en particulier, ainsi le « clustering » et le « sequence pattern mining » [Cheng et al., 2018] ne sont pas pris en compte et les règles d'association sont sélectionnées. Les règles d'association sont actuellement peu représentées dans les analyses de qualité [Köksal et al., 2011]. Nous utiliserons pour cela l'algorithme APriori [Witten et al., 2017].

Nous sélectionnons d'abord l'ensemble des individus de $T_{indicateurs}$ liés à C et qui sont alors stockés dans la table $T_{indicateurs_cible}$. Les données sont ensuite traitées pour les rendre interprétables. Pour cela, une table $T_{indicateurs_b}$ est créée. $T_{indicateurs_b}$ possède autant de colonnes que $T_{indicateurs_cible}$ de niveaux et est composée uniquement de valeurs booléennes. Le nom de chaque colonne

suivra le modèle suivant : « Ilot_Indicateur_Niveau ». Cette nomenclature sera utile pour la suite de l'étude des résultats.

Étape IV.2 : Sélection des paramètres de calcul

Si la table T_{indicateurs_b} est étudiée telle quelle avec l'algorithme APriori, un ensemble de règles sera retourné. Parmi ces règles, seuls les niveaux d'indicateurs les plus communs aux pièces produites sont présentés, car leur support est le plus important. Cependant, cela n'exprime pas pour autant de cause de mauvaise production. La génération des règles est donc guidée selon la méthodologie exprimée ci-après.

Dans un premier temps, les paramètres d'étude des règles sont définis. Une confiance invariable de 75 % est choisie pour toutes les études.

Dans un second temps, les supports sont variables en fonction des cas d'étude, autrement dit chaque produit a un support minimal et maximal différent. Afin de les obtenir, un raisonnement par ilot et adopté. Pour chaque ilot, le support minimum de chacun des indicateurs est calculé. Nous sélectionnons le minimum entre eux sur l'ilot puis sur la cartographie. Nous souhaitons pouvoir expliquer au moins 25 % de l'indicateur avec le plus petit support. Le support minimal est exprimé formellement par la formule suivante :

$$Support_{min} = \min_{Cartographie} (\min_{Domaine} (\min_{Indicateur} (Support))) \times 25\%$$

Un support maximal est également défini. La définition de ce support maximal permet d'éliminer les règles évidentes telles que la caractéristique la plus évaluée dans la table.

De manière semblable à la définition du support minimal, nous calculons les plus petits supports parmi chacun des niveaux des indicateurs sur un ilot. Parmi ceux-ci, le plus grand sur l'ilot est sélectionné. Enfin, le support maximal sera le plus petit de ces supports sur la cartographie. Ce niveau d'indicateur doit pouvoir être expliqué dans son ensemble. Le support le plus grand est exprimé par la formule suivante :

$$Support_{max} = \min_{Cartographie} (\max_{Domaine} (\min_{Indicateur} (Support)))$$

La Figure 4 présente de manière schématique la sélection du support maximal. La sélection par niveau y est détaillée.

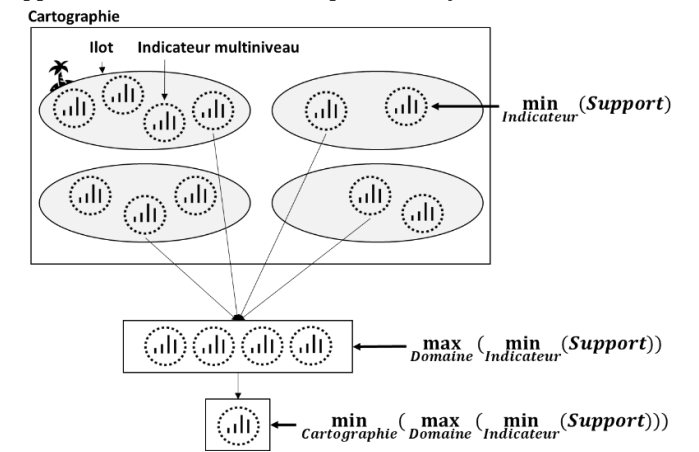


Figure 4 : Sélection du support maximal

Étape IV.3 : Résultats

Un ensemble de règles est généré, les règles intéressantes doivent être identifiées. Tout d'abord, l'ensemble des règles est affiché. Si l'une d'elles n'est pas évidente, elle est conservée dans une liste

que l'on notera E_{règles}. Nous effectuons ensuite une recherche par ilot. Pour cela, chaque règle contenant une variable de l'ilot, identifiable par la nomenclature définie dans la partie 3.5.1, dans les membres M₁ ou M₂ est triée dans un ensemble de règles portant le nom de l'ilot. Ces ensembles sont également étudiés, affichés et stockés dans E_{règles} le cas venu.

On réitère la manipulation pour les indicateurs spécifiques que l'on souhaite étudier.

Finalement, on affiche E_{règles} sous la forme d'un rapport. Cette liste de règles est ce qui aidera l'émission des décisions correctrices à prendre. Ces décisions ne sont pas générées automatiquement, car les solutions à un aléa peuvent être multiples selon le contexte de la production.

L'objectif de l'étude est de répondre à un des besoins liés aux enjeux de l'exploitation des données et en particulier sur l'analyse de la rentabilité d'un produit pour les PME. L'étude de la littérature ne fournit pas de modèle ou méthode générique applicable pour répondre à cet objectif. La méthodologie en 4 phases, présentée ci-dessus a donc été développée. Les résultats obtenus à la fin de ces 4 phases sont des rapports de production mettant en évidence les cooccurrences entre des caractéristiques et des aléas de production. La validation de cette méthodologie est présentée dans la partie suivante et s'appuie sur une étude de cas ayant eu lieu dans la PME APN.

4 CAS D'ETUDE

4.1 Contexte

La méthodologie présentée dans la partie 3 a été élaborée afin de répondre au besoin d'une entreprise d'usinage de précision : APN. Cette PME québécoise fournit des pièces pour l'industrie aéronautique. Elle est spécialisée dans la fabrication de petites pièces mécaniques complexes et utilise pour cela des machines à commande numérique.

4.2 Mise en œuvre de la démarche

4.2.1 Phase 1 : Cartographie des données

Étape I.1 : L'entreprise a un environnement de données se composant principalement de deux BDD, notées respectivement DB1 et DB2. DB1 est une base structurée par une entreprise d'ERP et DB2 est une base créée par APN et assimilable à une base de MES (Manufacturing Execution System). La Table 1 reprend les données générales caractérisant les bases. Dans un premier temps, il est nécessaire de réduire l'espace des BDD dans lequel on évolue. Pour cela, l'étude des différents flow charts, des rapports émis par le système et utilisé au quotidien, de même que l'étude du dictionnaire de tables de DB1 a été réalisé. À ces études papier, s'ajoutent les discussions avec les experts métiers. Les informations relatives aux réductions DB sont résumées dans la Table 1 suivante.

	DB1	DB2
Nombre de tables d'origine	152	373
Nombre de tables après réduction de l'espace	5	16

Étape I.2 : Les tables restantes sont organisées selon trois ilots. Les informations relatives (1) aux produits (2) aux processus et (3) aux opérateurs sont donc organisées distinctement les unes aux autres. Ces ilots se composent respectivement de X, Y et Z tables.

Ces ilots sont immédiatement reliés à la réduction de tables qui a été réalisée à l'étape précédente. En effet, le critère important d'APN pour l'OAD est de se centrer sur la qualité et rentabilité du produit. Le nombre de tables de la base contenant des informations relatives à ces paramètres est de 10 après réduction soit 62,5%. Si l'intérêt de l'entreprise avait été l'évaluation des relations fournisseurs, d'autres tables seraient apparues.

Étape I.3 : Les tables sont ensuite reliées les unes aux autres. Pour cela, on ignore 16 clés sur les 30 existantes pour n'en utiliser plus que 14. Ainsi le schéma des chemins sélectionnés n'a pas de cycles, ce qui signifie qu'il n'existe qu'une seule et unique manière de passer d'une table T_i à T_k . La Figure 5 représente schématiquement la cartographie ainsi que les cheminements entre les tables, soit la cartographie à la fin de l'étape I.3. On y observe l'absence de cycle.

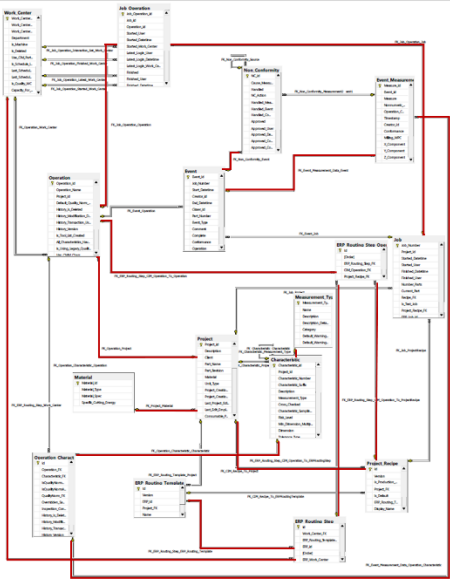


Figure 5 : Exemple de cartographie à la fin de l'étape I.3

Étape I.4 : Une liste de contraintes est éditée par table puis par ilot. L'ensemble des informations relatives aux tests de production ou de système ainsi que les données sur les services autres que la production sur machine (finition manuelle, logistique, achat) ne sont pas prises en compte pour l'étude.

Dans le cas de l'entreprise, la contrainte temporelle sur la plus jeune des bases utilisées est 2016. Cette réduction importante de l'espace est nécessaire, car la table concernée donne accès à 3 autres tables dont celle des machines utilisées lors de la production des pièces. Des contraintes supplémentaires sont également intégrées : on ne souhaite pas étudier (1) les données de produits qu'APN ne fabrique plus (2) les produits avec une production particulière.

4.2.2 Phase 2 : Mise en place d'indicateurs

Étape II.1 : La cartographie ainsi que les contraintes des tables sont définies, nous cherchons à mettre en place les indicateurs associés. Dans le cas d'APN, tous les ilots présents dans la première itération de la cartographie sont directement impliqués dans l'évaluation de critère de rentabilité. Des IG sont donc définis pour les trois ilots. Chaque IG est décomposé en plusieurs IL. Exemple : L'ilot de produits peut être mesuré selon (1) la matière (2) la criticité des dimensions du produit et (3) les Non Conformités (NC). Les NC se caractérisent concrètement par le pourcentage de pièces rebutées et le pourcentage de pièces

identifiées non conforme, mais ayant pu être retravaillé ou tout de même accepté par le client. Pour la suite de l'étude, on appelle « rebut » les pièces rebutées et NC l'ensemble pièces jugées non conformes à une étape de la production. Le pourcentage de pièces rebutées nécessite les informations brutes suivantes : l'état de la pièce à la fin de la production, le nombre de pièces par lot, la clé unique de la pièce pour être certain de ne pas compter plusieurs fois un rebut. La Figure 6 ci-dessous montre l'organisation des IG et IL par ilot. On y remarque que le nombre d'indicateurs est variable en fonction de l'ilot. De même qu'un IG peut être plus difficilement interprétable.

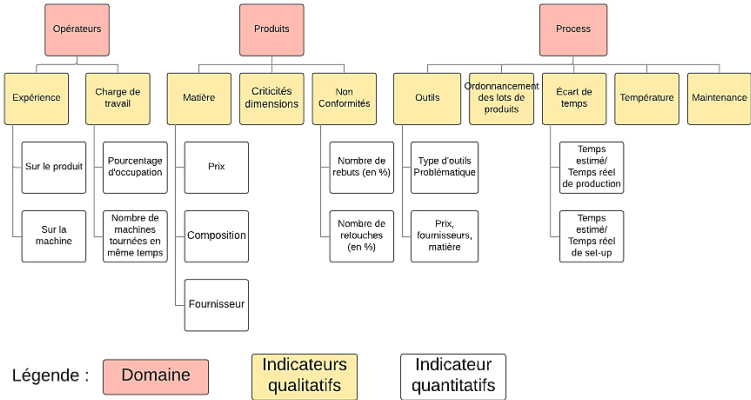


Figure 6 : Organisation des indicateurs par ilot

Étape II.2 : Les indicateurs ainsi sélectionnés sont intégrés à la cartographie selon les processus décrits Figures 2 et 3. Dans le cas des premiers indicateurs mis en place, seule une table est ajoutée à la cartographie de base.

4.2.3 Phase 3 : Utilisation quotidienne des KPIs

Étape III.1 : Les indicateurs sont définis et intégrés à la table $T_{indicateurs}$ et leurs tables inscrites à la cartographie. Tel qu'énoncé précédemment, un seul indicateur est sélectionné par ilot. La suite du cas d'étude est basée sur l'étude de l'indicateur des NC.

Note : Le nombre de défauts est adimensionné et ramené à un pourcentage du nombre total de pièces fabriquées.

Dans un premier temps, on rédige une liste de critères que doit respecter la visualisation. Leur élaboration vient des critères proposés par Parmenter (2015) ainsi que par les demandes de la compagnie. On obtient la liste des 5 contraintes suivantes, notées C_i :

- C1 - Voir la répartition des NC et des rebuts tout au long de la production.
- C2 - Voir l'historique de la production du produit.
- C3 - Comparer le pourcentage de NC d'une série à l'autre.
- C4 - Voir la taille de lot d'une série.
- C5 - Voir la plus grande série de NC pour le produit.

De cette liste de contraintes, la visualisation Figure 7 est obtenue. Sur cette figure, l'abscisse représente les pièces produites et l'ordonnée le pourcentage de NC d'un lot de pièces. On peut alors assimiler chaque ligne de pièces comme une série de pièces. Les numéros de série sont alors inscrits dans le coin supérieur gauche de la figure afin de pouvoir identifier les éventuels lots problématiques. L'intervalle de temps entre les séries n'est pas représenté sur ce graphique, le seul paramètre important dans ce cas est bien la succession d'une série à l'autre. Cette visualisation vérifie les contraintes C2, C3 et C4. La satisfaction de C1 vient de l'affichage des NC (barres verticales bleues) et des rebuts

(croix rouges). La contrainte C5 est vérifiée par l’affichage en légende de la plus grande série de pièces non conforme. Ce critère nous permet notamment de comprendre si les erreurs sont disséminées ou au contraire regroupées autour d’un même évènement. Dans le cas des NC, on définit l’élément de référence comme lié à l’historique de production. Si une série dépasse de 4 fois la moyenne antérieure des NC alors le produit est étudié.

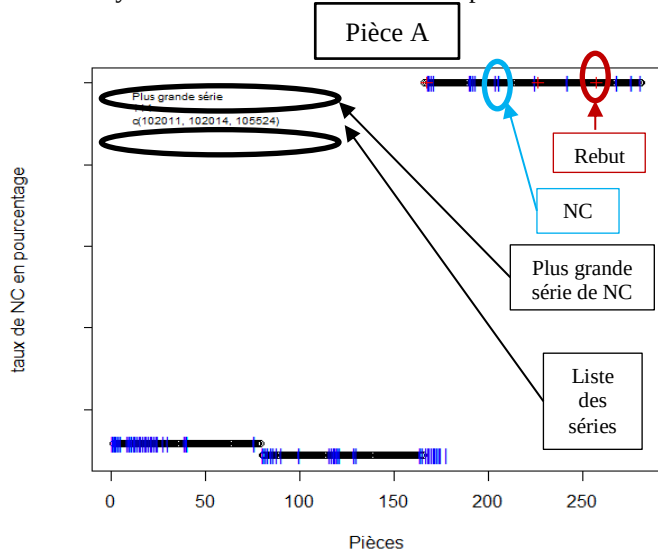


Figure 7 : Exemple de visualisation de l’indicateur des NC

Étape III.2 : La sélection d’un produit d’étude est justifiée par la visualisation d’une anomalie dans l’indicateur et l’absence de justification à cette anomalie. Dans ce cas, deux types d’anomalies sont distinguées (1) une série de pièces a un taux de NC anormalement haut et il y a dépassement de la référence (2) le taux des NC est instable entre les séries. Dans le cadre de (1) des patrons peuvent être identifiés et justifiés. Exemple : tous les problèmes sont arrivés au lancement, ou en fin de production, car une personne en formation a travaillé sur les premières pièces ou encore la machine est tombée en panne. L’identification de patrons peut rendre une étude plus poussée inutile et permet d’éviter des itérations de traitements inutiles.

Les Figures 7 et 8 donnent respectivement un exemple des anomalies de type (1) et (2). On observe bien sur la Figure 8 le phénomène en « dents de scie » sur les 6 dernières séries sans le dépassement de la référence, justifiant l’ajout de ce nouveau cas d’étude.

Pour la suite de l’étude, on se place dans le cas d’une anomalie de type (1) cf. Figure 7. Seule la série problématique est analysée. On note alors $T_{indicateurs_cible}$ l’extrait de la table $T_{indicateurs}$ ne contenant que les données relatives à la série. Une réduction de 335620 à 1198 est alors marquée.

4.2.4 Phase 4 : Analyse des données

Étape IV.1 : La transformation de $T_{indicateurs_cible}$ en $T_{indicateurs_b}$ correspond à la transformation d’une table de 1198 individus et 5 colonnes de flottants ou chaîne de caractères à une table de 1198 individus et 26 colonnes de booléens.

Étape IV.2 : Les résultats des calculs du minimum et du maximum d’individus sur chaque ilot sont résumés dans la table 2.

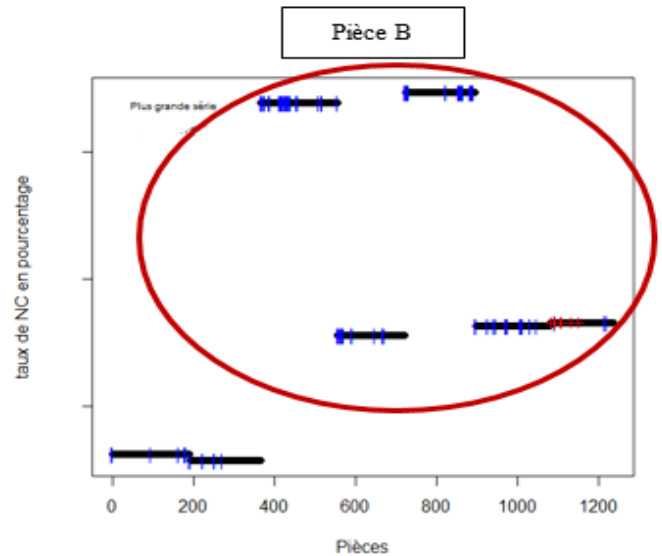


Figure 8 : Exemple d’anomalie sur la stabilité du processus

	Produit	Processus	Opérateur
Minimum	3	1198	43
Maximum	11	1198	410

Table 2 : Synthèse des paramètres de support par ilot

On observe que le minimum et maximum pour le processus sont égaux et qu’ils sont également égaux au nombre d’individus. Cela se traduit par le fait que pour la série étudiée les paramètres d’opérations étaient les mêmes : programme et numéro de projet identiques.

Étape IV.3 : L’utilisation de l’algorithme APriori sur $T_{indicateurs_b}$ nous donne un premier ensemble de 72 règles. Sur ces 72 règles, seules 4 ont un lift compris entre [2,9 et 38,6]. Celles-ci sont sélectionnées et insérées à $E_{règles}$. Aucune règle avec un lift plus petit que 0,9 n’est observé. On réitère une analyse par l’algorithme APriori sur chacun des ilots en faisant un tri sur les règles ayant un support supérieur à 2,9. L’étude sur l’ilot des opérateurs donne un nouvel ensemble de 12 règles présentées dans la Figure 9. De cette visualisation, on peut en déduire l’opérateur qui a réalisé le plus de NC qui ne sont pas des rebuts « Is_Scrap-0 », mais aussi celui qui en a réalisé le plus de rebuts « Is_Scrap-1 ». On identifie également la nature de la cote à l’origine des rebuts et aussi le numéro de projet associé.

La méthodologie s’achève sur la présentation de $E_{règles}$. Les décisions sont prises par les responsables de l’entreprise.

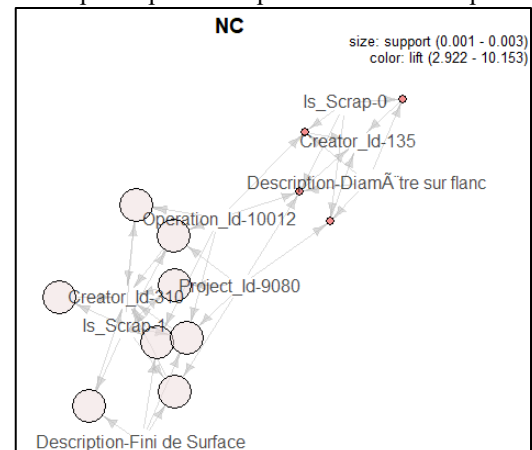


Figure 9 : Résultat des règles d’associations sur l’ilot « Opérateur »

La méthodologie s'achève sur l'émission d'un rapport sur la production d'une série de produits. Ce rapport concerne dans notre cas les opérateurs, le processus théorique et les caractéristiques de produits. Le rapport généré porte sur la mise en évidence de cooccurrences d'événements lors de la production et ainsi la liaison entre aléas et paramètres de production.

5 CONCLUSION

L'enjeu étudié était de proposer une méthodologie pour la conception d'un outil de traitement de données, afin d'identifier des facteurs et schémas responsables de la baisse de productivité pour une PME. Après le constat d'une absence de méthode générique dans la littérature nous proposons pour répondre à cet enjeu une méthodologie répliquable de 4 phases. Celles-ci sont (1) la mise en place d'une cartographie, (2) la mise en place d'indicateurs, (3) l'utilisation de KPIs et enfin (4) l'analyse des données. Le résultat de ce cheminement est un rapport liant aléas et caractéristiques de production. Cette méthodologie a été validée par une étude de cas.

Les résultats obtenus permettent l'identification de paramètres de production impactant la rentabilité du produit. Cette identification génère alors une prise de décision humaine sur la production.

Cependant, les rapports générés ne suggèrent pas et ne guident pas les dirigeants dans leur prise de décisions. L'OAD serait alors plus performant si la prise de décision était davantage assistée, voire automatisée. Une amélioration de l'interface homme-machine couplée à l'utilisation de nouvelles fonctions du data mining est selon nous des pistes exploitables afin de compléter l'outil.

La non-exhaustivité des informations relatives à la production pour la génération du rapport est une limite de la méthodologie de cette étude. La prise en considération de l'environnement de production soit le critère (5) de [Schuh et al., 2015] et la maîtrise des informations collectées est un point qui serait important de développer selon nous. Cependant, l'étude dans un environnement simplifié nous a permis d'obtenir de premiers résultats significatifs.

6 REMERCIEMENTS

Merci à l'entreprise APN d'avoir participé à l'élaboration de cette méthodologie ainsi que MITACS pour la subvention IT12642.

7 REFERENCES

Buddhakulsomsiri, J., Siradeghyan, Y., Zakarian, A., Li, X., (2006) Association rule-generation algorithm for mining automotive warranty data. *International Journal of Production Research*, pp. 2749-2770.

Cheng, Y., Chen, K., Sun, H., Zhang, Y. Tao, F., (2018), Data and knowledge mining with big data towards smart production, *Journal of Industrial Information Integration*, 9, pp. 1-13.

Choudhary, A. K., Harding, J. A., Tiwari, M. K., (2009) Data mining in manufacturing: a review based on the kind of knowledge, *Journal of Intelligent Manufacturing*, 20(5), pp. 501 – 521.

Danjou, C., Rivest, L., Pellerin, R., (2017), Le passage au numérique, Cefrio : Montréal (Canada).

Dobos, P., Tamás, P., Illés, B., Balogh, R., (2018) Application possibilities of the Big Data concept in Industry 4.0, *IOP Conference Series: Materials Science and Engineering*, 448.

Drath, R., Horch, A., (2014), Industrie 4.0 – hit or hype? *IEEE*

Industrial Electronics Magazine, 8(2): pp. 56-58.

Grabot, B., (1998) Objective satisfaction assessment using neural nets for balancing multiple objectives, *International Journal of Production Research*, 36(9), pp. 2377-2395.

Halaškal, M., Šperka, R., (2018) Process Mining - The Enhancement of Elements of Industry 4.0, *4th International Conference on Computer and Information Sciences (ICCOINS)*, pp. 1-6.

Hély, V., Antoni, J.-P., (2019) Combining indicators for decision making in planning issues: A theoretical approach to perform sustainability assessment, *Sustainable Cities and Society*, 44, pp. 844-854.

Kagermann, H. (2015). Change through digitization—value creation in the age of industry 4.0. *Management of Permanent Change*, Springer Science Business Media, pp. 23-45.

Kimball, R., Ross, M. (2013) The data warehouse Toolkit 3rd Ed., Wiley: Indiana (États-Unis).

Köksal, G. Batmaz, I., Testik, M.C., (2011) A review of data mining applications for quality improvement in manufacturing industry, *Expert Systems with Applications*, 38, pp. 13448–13467.

Lemberger, P., Batty, M., Morel M., Raffaelli, J. L. (2016) Big Data et Machine Learning 2nd Ed., Dunod: Paris (France).

Linoff, G.S, Berry, M.J.A., (2011), Data mining Techniques for marketing, sales and customer relationship management, 3rd Ed., Wiley: Indiana (États-Unis).

Nemeth, M., Peterkova, A., (2018) Proposal of data acquisition method for industrial processes in automotive industry for data analysis according to Industry 4.0, *IEEE International Conference on Intelligent Engineering Systems (22nd)* : Las Palmas de Gran Canaria, Spain 21-23 Juin.

Packianathera, M.S., Davies, A., Harradana, S., Somanb, S., White, J., (2017) Data mining techniques applied to a manufacturing SME, *CIRP Conference on Intelligent Computation in Manufacturing Engineering 10th*, pp. 123-128.

Parmenter, D., (2015) Key Performance Indicators: Developing, Implementing, and Using Winning KPIs: Wiley: Indiana (États-Unis).

Rissoan, R., Jouin, R., (2018) La boîte à outils de la stratégie big data, Dunod : Saint Étienne (France).

Schuh, G., Gartzten, T., Rodenhauser, T., Marks, A., (2015) Promoting work-based learning through industry 4.0, *5th Conference on Learning Factories 2015*, pp. 82-87.

Tamim, N., Laboureur, D., Rashid, A., Sam M., (2018), Developing leading indicators-based decision support algorithms and probabilistic models using bayesian network to predict kicks while Drilling, *Process Safety and Environmental Protection*, pp. 239-246.

Vitalii, I., Yiming R.J.T., Joachim, V., (2018) *Advances in Design, Simulation and Manufacturing*, Springer: Switzerland.

Witten, I., Frank, E., Hall, M., Pal, C., (2017) Data Mining Practical Machine Learning Tools and Techniques, *Data Management Systems Series*, 4th Ed., Morgan Kaufmann.

Yahyaoui, H., (2015) Méthode d'analyse de données pour le diagnostic a posteriori de défauts de production - Application au secteur de la microélectronique. *Ecole Nationale Supérieure des Mines de Saint-Etienne (EMSE)*, 0795. Saint-Etienne (France).