



ÉCOLE

CENTRALE LYON

Étude Bibliographique :
Machine Learning en Chaîne
d'Approvisionnement

Réalisé par :

Melis ULGENALP

Sommaire

Liste des Figures.....	3
Liste des Tableaux.....	4
Introduction.....	5
La Chaîne d'Approvisionnement.....	6
La Gestion de la Chaîne d'Approvisionnement	7
Machine Learning	8
Applications Techniques de Machine Learning	8
I. Prévion de la demande dans la chaîne d'approvisionnement	8
II. Prévion du risque dans la chaîne d'approvisionnement	15
III. Une approche d'apprentissage machine supervisée pour la simulation pilotée par les données de la sélection de fournisseurs résilients dans la fabrication numérique.....	23
Conclusion	28
Bibliographie	29

Liste des Figures

Figure 1: Un exemple de chaîne d'approvisionnement de base	6
Figure 2: Les étapes de la chaîne d'approvisionnement	7
Figure 3: La méthode de réseau de neurones avec feed-forward error back-propagation	11
Figure 4: La méthode des réseaux de neurones récurrents.....	12
Figure 5: La modèle d'un partenaire.....	13
Figure 6: Le cadre proposé dans l'article	17
Figure 7: La variance de score de prédiction pour chaque sous-ensemble de 33 caractéristiques.	21
Figure 8: Les flux d'informations et de matières dans le modèle de simulation.....	24
Figure 9: Le modèle de sélection du fournisseur par apprentissage supervisé de la machine.....	25
Figure 10: L'algorithme de Hybride A.	27
Figure 11: L'algorithme de Hybride B.	27
Figure 12: Les fiabilités des méthodes différentes	28

Liste des Tableaux

Tableau 1: Les méthodes d'estimation de la demande utilisées.....	9
Tableau 2: Comparaison de la performance (Err. Moy.) des techniques de prévision sur l'ensemble des données de simulation.....	14
Tableau 3: Comparaison de la performance (Err. Moy.) des techniques de prévision sur l'ensemble de données sur les fonderies.	14
Tableau 4: Les sources de données externes et internes	18
Tableau 5: Un résumé des meilleurs scores de prédiction pour différents classificateurs et métriques	22

Introduction

De nos jours, avec l'augmentation de la puissance des ordinateurs et des sources de données, l'intérêt pour l'apprentissage machine dans les entreprises augmente (Baryannis, Dani, & Antoniou, 2019). De plus, l'apprentissage machine est utilisé pour améliorer davantage les chaînes d'approvisionnement afin d'obtenir d'en obtenir des plus rentables et efficaces. Ainsi, il convient d'abord de définir les termes suivants : « chaîne d'approvisionnement », « gestion de la chaîne d'approvisionnement » et « apprentissage machine ». Ensuite, nous analyserons trois articles écrits sur ce sujet. Le premier article traite de la prévision de la demande dans la chaîne d'approvisionnement, le deuxième aborde la prévision du risque dans la chaîne d'approvisionnement. Enfin, le troisième article donne une approche d'apprentissage machine supervisée pour la simulation pilotée par les données de la sélection de fournisseurs résilients dans la fabrication numérique. C'est donc à travers ces articles que nous examinerons trois différentes applications de l'apprentissage machine dans la chaîne d'approvisionnement. Notamment, avec le deuxième et le troisième article, nous explorerons de plus près la façon dont l'apprentissage machine peut être utilisé avec d'autres technologies comme la simulation.

La Chaîne d'Approvisionnement

Selon le Conseil de la Chaîne d'Approvisionnement (Supply Chain Council), les définitions de ce terme sont les suivantes (CSCMP Supply Chain Management Definitions and Glossary, 2013) :

1) En commençant par les matières premières non transformées et en aboutissant au client final utilisant les produits finis, la chaîne d'approvisionnement relie de nombreuses entreprises.

2) Les échanges matériels et informationnels dans le processus logistique allant de l'acquisition des matières premières à la livraison des produits finis à l'utilisateur final. Tous les fournisseurs, prestataires de services et clients sont des maillons de la chaîne d'approvisionnement.

Christopher (1994) définit la chaîne d'approvisionnement comme « un réseau d'organisations impliquées, par des liens en amont et en aval, dans les différents processus et activités générant de la valeur sous la forme de produits et de services entre les mains du client final ».

Un exemple de chaîne d'approvisionnement de base est présenté à la figure suivante (Chopra & Meindl, 2001) :



Figure 1: Un exemple de chaîne d'approvisionnement de base

La Gestion de la Chaîne d'Approvisionnement

La définition donnée par le Conseil des professionnels de la gestion de la chaîne d'approvisionnement (Council of Supply Chain Management Professionals-CSCMP) est la suivante (CSCMP Supply Chain Management Definitions and Glossary, 2013) :

« La gestion de la chaîne logistique englobe la planification et la gestion de toutes les activités liées à l'approvisionnement, aux achats, à la conversion et à toutes les activités de gestion logistique. De manière importante, cela inclut également la coordination et la collaboration avec les partenaires de canal, qui peuvent être des fournisseurs, des intermédiaires, des fournisseurs de services tiers et des clients. En substance, la gestion de la chaîne d'approvisionnement intègre la gestion de l'offre et de la demande dans et entre les entreprises. La gestion de la chaîne logistique est une fonction intégratrice dont la responsabilité première est de lier les principales fonctions et processus d'entreprise au sein d'une entreprise à l'autre et au sein d'une entreprise dans un modèle opérationnel cohérent et très performant. Il inclut toutes les activités de gestion logistique mentionnées ci-dessus, ainsi que les opérations de fabrication. Il pilote la coordination des processus et des activités avec et entre le marketing, les ventes, la conception de produits, les finances et les technologies de l'information. »

Selon Chopra et Meindl (2001), la gestion de la chaîne d'approvisionnement implique la gestion des flux entre les étapes d'une chaîne d'approvisionnement afin de maximiser la rentabilité totale.

Dans la figure ci-dessus, nous pouvons observer les étapes de la chaîne d'approvisionnement (Christopher, 1994) :

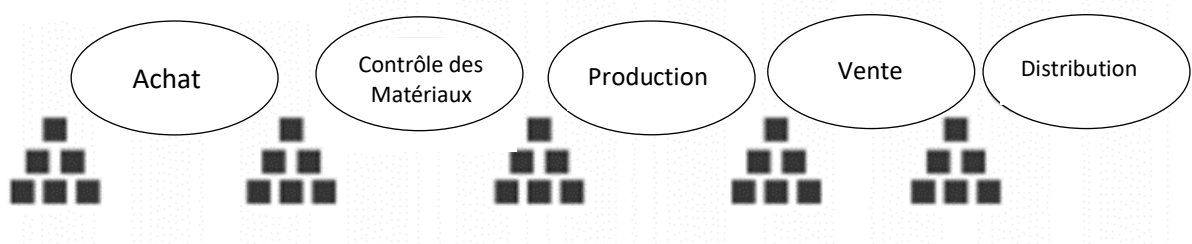


Figure 2: Les étapes de la chaîne d'approvisionnement

Machine Learning

Machine Learning (ML) décrit un système qui apprend automatiquement des programmes à partir de données. Au lieu de créer manuellement des programmes, un modèle ML sera formé avec un ensemble de données existant. Ensuite, le modèle ML peut effectuer des tâches apprises sur de nouvelles données. L'acquisition de connaissances à partir de données nécessite un ensemble d'exemples de cas et de fonctions d'entrée pertinentes. Cela implique que ML nécessite une taille et une qualité suffisantes des données pouvant être utilisées pour former le modèle.

Applications Techniques de Machine Learning

I. Prévision de la demande dans la chaîne d'approvisionnement

Dans cette partie d'étude, nous examinerons l'article suivant : (Carbonneau, Laframboise, & Vahidov, 2008).

Le but de cet article est d'étudier l'applicabilité des techniques avancées d'apprentissage automatique telles que les réseaux de neurones, les réseaux de neurones récurrents et la machine à vecteur de support pour estimer les demandes faussées à la fin de la chaîne d'approvisionnement, c'est-à-dire l'effet coup de fouet. Dans cette recherche, les techniques avancées d'apprentissage automatique sont comparées aux méthodes traditionnelles d'estimation de la demande et la faisabilité des techniques avancées d'apprentissage automatique est étudiée. Les méthodes traditionnelles d'estimation de la demande utilisées pour la comparaison dans la recherche sont les suivantes : prévisions naïves, moyenne mobile, régression linéaire et modèles de séries chronologiques.

Bien que la diminution des coûts de la chaîne d'approvisionnement étendue ne soit pas l'objectif principal de cette recherche, on estime que les coûts de la chaîne d'approvisionnement seront généralement réduits en augmentant la précision des prévisions de la demande dans la chaîne d'approvisionnement. Cette diminution des coûts sera due à la baisse des coûts d'inventaires et à la satisfaction accrue des clients. Afin d'augmenter la précision des prévisions de la demande dans la chaîne d'approvisionnement, l'effet coup de fouet doit être interprété. Un léger retard dans la chaîne d'approvisionnement a un impact

imprévisible à mesure que la chaîne d'approvisionnement progresse par étapes. Ce phénomène pourrait être expliqué en termes de théorie du chaos, où une petite variation de l'entrée pourrait entraîner de grandes, apparemment aléatoires, comportements en sortie (Kullback, 1968).

Dans cette section, nous présenterons les méthodes d'estimation de la demande utilisées dans l'article.

Méthodes Traditionnelles	Méthodes d'Apprentissage Automatique Non Linéaires
<ul style="list-style-type: none"> • prévision naïve, • moyenne, • moyenne mobile, • trend • régression linéaire multiple 	<ul style="list-style-type: none"> • réseau de neurones, • réseau de neurones récurrents, • machines à vecteurs de support

Tableau 1: Les méthodes d'estimation de la demande utilisées

Méthodes Traditionnelles :

Prévision Naïve : Technique d'estimation dans laquelle les chiffres réels de la dernière période sont utilisés comme prévisions pour cette période, sans les ajuster, ni tenter d'établir des facteurs de causalité. Il n'est utilisé qu'à des fins de comparaison avec les prévisions générées par les meilleures techniques (sophistiquées) (Business Dictionary, 2019).

Prévision Moyenne Mobile : Il utilise la moyenne d'un nombre défini de périodes précédentes, comme la demande future prévue. La formule de la moyenne mobile (MM) est :

$$MM = \frac{A_1 + A_2 + \dots + A_n}{n}$$

A est la moyenne de la période n et **n** est le nombre de périodes (Hayes, 2019).

Prévision Basée Sur le Trend: Il est basé sur un modèle de régression simple qui définit le temps comme variable indépendante et tente de prévoir la demande en fonction du temps.

Régression Linéaire Multiple (RLM): est une technique statistique qui utilise plusieurs variables explicatives pour prédire le résultat d'une variable de réponse. Le but de la régression linéaire multiple est de modéliser la relation linéaire entre les variables explicatives (indépendantes) et la réponse (dépendante) (Kenton, 2019). La formule de la régression linéaire multiple est :

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon$$

y_i est une variable dépendante, x_i est une variable explicative, β_0 est un terme constant, β_p est le coefficient de pente pour chaque variable explicative, et ε est le terme d'erreur du modèle.

Méthodes d'Apprentissage Automatique Non Linéaires

Réseau de Neurones (RN): Bien qu'il existe de nombreux types de méthodes de réseaux de neurones artificiels, la méthode *feed-forward error back-propagation* est utilisée dans cette recherche. « Dans ces réseaux, neurones sont organisés en couches de telle sorte que les signaux de sortie des neurones d'une couche donnée sont transmis à tous les neurones de la couche suivante. Ainsi, le flux d'activations neurales va dans une seule direction, couche par couche. Le plus petit nombre de couches est deux, à savoir l'entrée et couches de sortie. Plus de couches, appelées couches cachées, pourrait être ajouté entre l'entrée et la sortie couche. La fonction des couches cachées est d'augmenter la puissance de calcul des réseaux de neurones » (Carbonneau, Laframboise, & Vahidov, 2008).

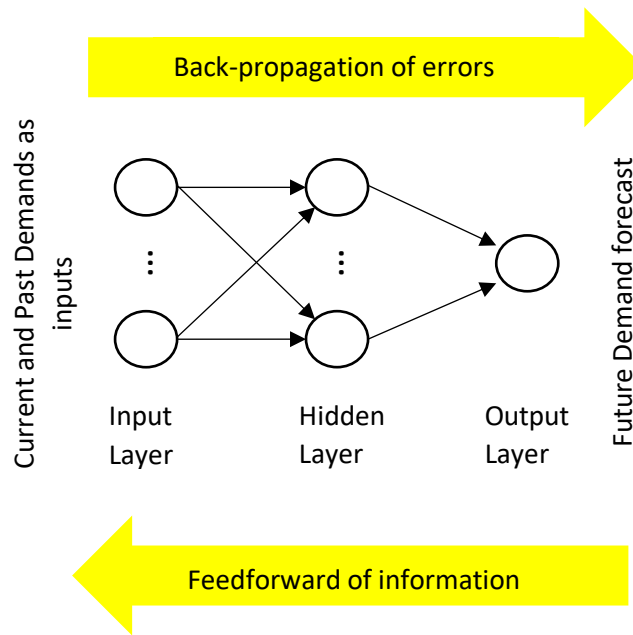


Figure 3: La méthode de réseau de neurones avec feed-forward error back-propagation

Réseaux de Neurones Récurrents (RNR) : « Les réseaux neuronaux récurrents permettent aux signaux de sortie de certains de leurs neurones de refluer et servent d'entrées pour les neurones de la même couche ou ceux des couches précédentes. Le RNR est un outil puissant pour de nombreux problèmes complexes, en particulier les données de séries chronologiques sont impliquées. La méthode de formation appelé « *back-propagation* dans le temps » pourrait être RNR sur un ensemble d'entraînement donné (Werbos, 1990). La Figure 4 montre schématiquement la structure du RNR pour la prévision de la demande de la chaîne d'approvisionnement problèmes. Comme on peut le constater sur la Figure 4, seule la sortie (output) des neurones cachés est utilisée pour servir d'entrée (input) aux neurones de la même couche. » (Carbonneau, Laframboise, & Vahidov, 2008)

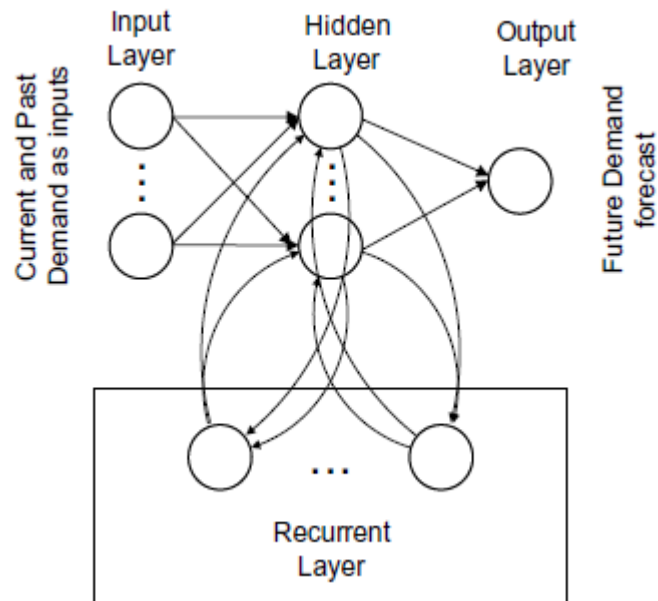


Figure 4: La méthode des réseaux de neurones récurrents

Machines à Vecteurs de Support : « Les machines à vecteurs de support (SVM) sont basées sur le principe de minimisation des risques structurels de la théorie de l'apprentissage statistique (Vapnik, 1995) comme opposé au principe empirique de minimisation des risques sur quels réseaux de neurones et régression linéaire sont basés. L'objectif de la structure minimisation des risques invisible, opposé à RN et RLM, ce qui minimisait l'erreur. Ils projettent les données dans une dimension supérieure l'espace et maximiser les marges entre ou en minimisant la marge d'erreur pour la régression. Les marges sont « soft », ce qui signifie qu'une solution peut être des exemples contradictoires dans l'ensemble d'entraînement. Le problème est formulé comme un convexe d'optimisation sans minimum local, fournissant ainsi une solution unique par opposition à la *back-propagation* réseaux de neurones, qui peuvent avoir plusieurs minima et, par conséquent, ne peut garantir que l'erreur minimale sera atteinte » (Carbonneau, Laframboise, & Vahidov, 2008).

Data

Deux sources de données de demande sont utilisées dans la recherche. La première source de données est une simulation de la chaîne d'approvisionnement étendue ; le deuxième se compose des données mensuelles sur les ventes d'une fonderie de Statistics Canada.

Cette simulation de chaîne d'approvisionnement étendue a été développée dans MATLAB. Elle se compose de quatre lots, de la production du produit au client final. Basé sur cette chaîne d'approvisionnement étendue modèle, le traitement du signal de demande simulé est introduit comme source de distorsion de la demande. Le traitement du signal de demande est modélisé avec la régression linéaire simple qui calcule la tendance au cours des 10 derniers jours, et qui est ensuite utilisé pour prévoir la demande en 2 jours. Dans la simulation, chaque partenaire de la chaîne d'approvisionnement est modélisé de la même manière. Les tâches d'un partenaire dans la simulation sont les suivantes : prévision de la demande et calcul des bons de commande, acceptation des commandes, acceptation des marchandises et livraison des marchandises. Le diagramme suivant montre la modèle d'un partenaire.

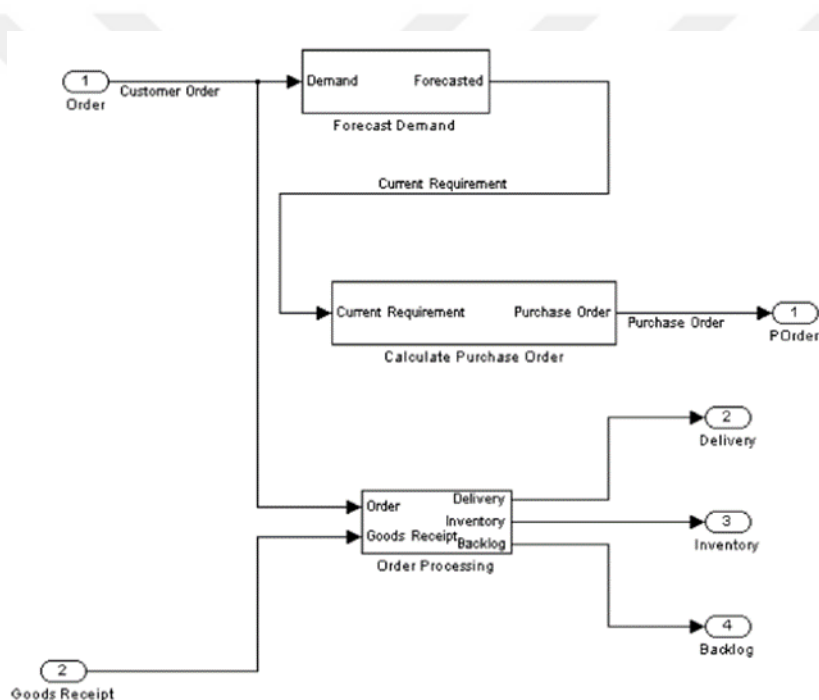


Figure 5: La modélisation d'un partenaire

Les recherches de Carbonneau, Laframboise et Vahidov ont montré que les applications ML permettent d'obtenir des prévisions de demande plus précises que les méthodes traditionnelles comme la prévision naïve, moyenne, moyenne mobile, trend et régression linéaire multiple (Kullback, 1968). Après préparation des données, des méthodes d'apprentissage automatique non linéaires (réseau de neurones, réseau de neurones récurrents et machines à vecteurs de support) et des méthodes traditionnelles ont été appliquées aux données. Ensuite, ces deux types de méthodes ont été comparés.

Nous procédons ensuite à la comparaison des méthodes. Les résultats des expériences suggèrent que les réseaux de neurones récurrents et les machines à vecteurs de support ont été les techniques de prévision les plus précises. Les moyennes mobiles, les prévisions naïves et le trend ont été parmi les moins performants. Cependant, l'analyse statistique des résultats montre qu'il n'y avait pas de différences statistiquement significatives en termes de précision des prévisions entre réseau de neurones récurrents, machines à vecteurs de support, réseau de neurones et régression linéaire multiple. (Carbonneau, Laframboise, & Vahidov, 2008) La comparaison des résultats des différentes prévisions en paramètres erreur moyenne (mean average error) et déviation standard pour les données de simulation et pour l'ensemble de données fonderies sont présentées aux tableaux ci-dessous (Carbonneau, Laframboise, & Vahidov, 2008) :

Technique de Prévision	L'ensemble de test		L'ensemble de formation	
	Err. Moy.	Dév.Std.	Err. Moy.	Dév.Std.
Réseau de neurones récurrents	447.72	328.23	461.66	350.35
Machines à vecteurs de support	453.04	341.88	449.32	365.01
Régression linéaire multiple	453.22	343.65	464.62	375.62
Réseau de neurones	455.41	354.40	471.03	383.29
Prévision naïve	520.53	407.29	536.45	435.05
Moyenne mobile	526.61	370.35	558.82	400.11
Trend	618.02	487.42	674.68	490.50

Tableau 2: Comparaison de la performance (Err. Moy.) des techniques de prévision sur l'ensemble des données de simulation

Les valeurs ci-dessus décrivent une certaine corrélation entre l'erreur moyenne et la déviation standard pour l'ensemble de données de simulation (sauf dans le cas de la moyenne mobile). Généralement, la déviation standard augmente avec l'augmentation de l'erreur moyenne.

Technique de Prévision	L'ensemble de test		L'ensemble de formation	
	Dév.Std.	Err. Moy.	Dév.Std.	Err. Moy.
Réseau de neurones récurrents	20.352	16.203	15.521	12.334
Machines à vecteurs de support	20.485	17.304	3.665	3.722
Régression linéaire multiple	21.396	19.705	15.007	15.041
Réseau de neurones	25.260	19.733	12.855	12.057
Moyenne mobile	25.481	19.253	18.205	13.028
Trend	27.323	24.198	17.995	17.292
Prévision naïve	32.591	23.485	20.263	17.380

Tableau 3: Comparaison de la performance (Err. Moy.) des techniques de prévision sur l'ensemble de données sur les fonderies.

Pour l'ensemble de donnée fonderies, on ne peut pas observer directement cette corrélation. Bien qu'il y ait un trend, on ne peut pas généraliser cette situation.

II. Prévision du risque dans la chaîne d'approvisionnement

Dans cette partie d'étude, nous examinerons l'article suivant (Baryannis, Dani, & Antoniou, 2019) :

La définition de la gestion des risques de la chaîne d'approvisionnement telle que définie dans l'article est la suivante :

« englobe les efforts de collaboration et de coordination de toutes les parties impliquées dans une chaîne d'approvisionnement pour identifier, évaluer, atténuer et surveiller les risques dans le but de réduire la vulnérabilité et d'accroître la robustesse et la résilience de la chaîne d'approvisionnement, en assurant la rentabilité et la continuité » (Baryannis, Validi, Dani, & Antoniou, 2018).

Selon l'article, les pratiques de gestion des risques de la chaîne d'approvisionnement peuvent être divisées en trois groupes principaux : (1) les techniques d'analyse décisionnelle à critères multiples ; (2) la modélisation mathématique et l'optimisation ; et (3) les techniques d'intelligence artificielle. Le troisième groupe a reçu moins d'attention en recherche que les deux autres groupes. Cependant, cela change en raison de l'augmentation de la puissance de l'ordinateur et de l'accès accru aux ressources d'information. Aujourd'hui, l'utilisation de l'intelligence artificielle augmente dans les domaines de la gestion de la chaîne d'approvisionnement et de la gestion des risques de la chaîne d'approvisionnement (Baryannis, Dani, & Antoniou, 2019).

L'intelligence artificielle dans la gestion des risques de la chaîne d'approvisionnement est utilisée pour les tâches l'identification des risques, la prévision des risques, l'évaluation des risques et réponses.

Cet article traite de l'utilisation des techniques d'apprentissage automatique dans la gestion des risques de la chaîne d'approvisionnement. Dans cette étude, les techniques d'apprentissage automatique sont évaluées avec des niveaux d'interprétabilité. Dans cette

étude, le « framework » proposé a été appliqué à une étude de cas utilisant deux techniques d'apprentissage automatique différentes. Les techniques d'apprentissage automatique utilisées sont la machine à vecteurs de support et l'apprentissage par arbre de décision. Dans cette étude de cas, l'estimation du risque est faite par classification. Les classificateurs sont formés pour déterminer s'il existe un risque de retard de livraison.

Contributions de cet article à la littérature : un cadre (framework) de prévision du risque fondé sur des données, une mise en œuvre du cadre proposé, la quantification du compromis entre l'interprétabilité et la performance de la prévision au moyen d'expériences pour l'étude de cas particulière (Baryannis, Dani, & Antoniou, 2019).

Tout d'abord, nous examinerons le cadre proposé dans l'article. Nous considérerons ensuite la mise en œuvre de ce cadre dans l'étude de cas.

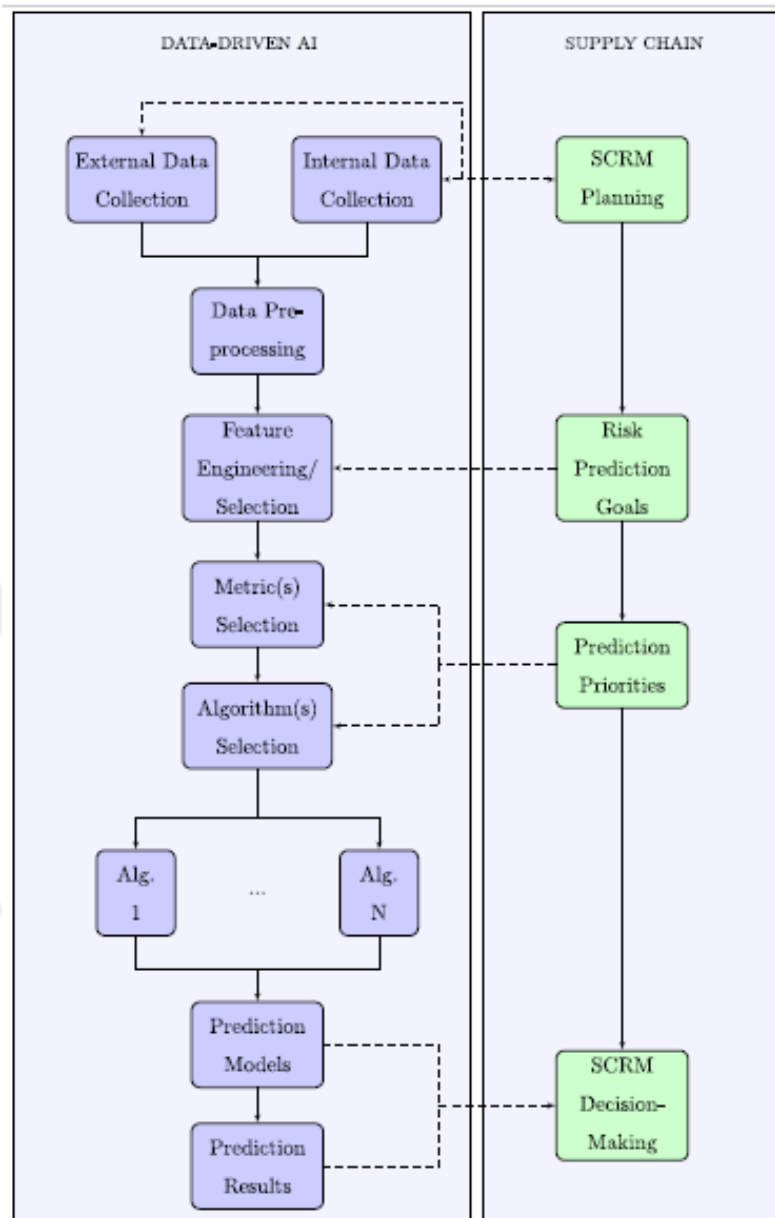


Figure 6: Le cadre proposé dans l'article

Le cadre développé pour l'application des techniques d'apprentissage automatique dans la gestion des risques de la chaîne d'approvisionnement est le suivant : selon ce cadre, nous devons d'abord établir un plan de gestion des risques pour la chaîne d'approvisionnement. Après avoir créé ce plan, les données pertinentes doivent être collectées. Ces données peuvent être regroupées en deux groupes, externes et internes. Le tableau suivant répertorie les sources de données externes et internes (Baryannis, Dani, & Antoniou, 2019) :

Données Externes	Données Internes
<ul style="list-style-type: none"> • Articles • Rapports météo • Activité des médias sociaux • Politiques nationales et internationales 	<ul style="list-style-type: none"> • Achats • Production • Dossiers de livraison et de vente • Informations sur le GPS et le capteur de conteneur • Données sur les finances de l'entreprise • Données sur les ressources humaines de l'entreprise

Tableau 4: Les sources de données externes et internes

Une fois les sources de données identifiées, les données doivent être collectées rétrospectivement. Après cette étape, les objectifs et les priorités de la chaîne d'approvisionnement doivent être déterminés. Ces cibles peuvent également être déterminées en fonction de la contrainte de données disponible. Les cibles que nous nous fixons auront un impact direct sur la sélection de l'algorithme d'intelligence artificielle à utiliser dans l'estimation des risques. À ce stade, la communication entre les experts de la chaîne d'approvisionnement et les experts de l'intelligence artificielle est cruciale pour la réussite du projet.

Data preprocessing a un impact direct sur le succès de l'estimation des risques. Le prétraitement des données englobe un large éventail de procédures, telles que la garantie que les données sont dans un format lisible par la machine, le nettoyage des données pour éliminer les entrées incohérentes ou incomplètes, le traitement des valeurs inconnues dans les données et la mise à l'échelle, la transformation, la normalisation ou la discrétisation des données (Baryannis, Dani, & Antoniou, 2019).

Feature engineering consiste à ajouter à l'ensemble de caractéristiques extraites initialement en combinant des variables et/ou des caractéristiques pour créer de nouvelles caractéristiques qui peuvent augmenter la puissance prédictive. Le but ultime de ces processus est de transformer un simple ensemble de valeurs dans un ensemble de données

en un ensemble significatif de caractéristiques qui a le potentiel d'aider à atteindre les objectifs de prédiction des risques définis précédemment (Baryannis, Dani, & Antoniou, 2019).

Comme le montre le diagramme, la sélection métrique et la sélection de l'algorithme dépendent de la définition d'objectifs de prédiction des risques.

L'algorithme sélectionné peut produire les sorties suivantes,

- True Positive/Vrai positif (TP) : « le modèle prédit correctement la manifestation du risque. Cela permet à la chaîne d'approvisionnement de se préparer à un tel résultat suffisamment tôt pour en atténuer les conséquences et toute ressource dépensée à cette fin est justifiée. Par conséquent, les vrais positifs doivent être maximisés. »
- True Negative/Vrai négatif (TN) : « le modèle prédit correctement que le risque ne se manifestera pas. Ces derniers doivent également être maximisés, mais leur importance peut être inférieure à celle des TP. »
- False Positive/Faux positif (FP) : « le modèle prédit incorrectement que le risque se manifestera, ce qui signifie que des ressources d'atténuation ont été dépensées inutilement. Si l'on accorde une priorité plus élevée à l'économie de ressources qu'à la saisie de tous les cas probables de manifestation du risque, alors une mesure qui accorde plus d'importance aux FP qu'aux TP. »
- False Negative/Faux négatif (FN) : « le modèle prédit incorrectement que le risque ne se manifestera pas. Il s'agit peut-être du résultat le plus indésirable, puisqu'essentiellement, l'objectif de prédiction du risque n'est pas atteint. Dans certains cas, la priorité peut être de minimiser les FN, même au détriment d'une augmentation des PF » (Baryannis, Dani, & Antoniou, 2019).

La définition de la métrique de « précision », qui est fréquemment utilisée dans ce type de recherche, est la suivante (Baryannis, Dani, & Antoniou, 2019) :

$$\frac{TP + TN}{PP + PN} \text{ avec } PP = TP + FP \text{ et } PN = TN + FN$$

ETUDE DE CAS

Dans cette étude de cas, on va examiner l'applicabilité du framework proposé en chaîne d'approvisionnement de la fabrication aérospatiale. Cette entreprise a une chaîne d'approvisionnement à plusieurs niveaux avec plusieurs partenaires européens et asiatiques. On a un jeu de données qui contient des informations de 50 000 livraisons du produits, les informations d'une période de six ans (2011-2016). Pour chaque livraison, on dispose des informations suivantes :

- Fournisseur de niveau 1 (Tier 1 supplier) : nom et ID
- Fournisseurs de niveau 2 : nom de fournisseur et ID
- Produits : numéro de pièce et description de pièce ; prix par unité
- Commandes : numéro de commande, date et numéro de ligne ; quantité commandée ; date d'échéance ; date de livraison demandée à l'origine ; date de livraison actuellement acceptée
- Livraisons : date de réception ; quantité reçue ; quantité refusé ; statut de livraison de la ligne de commande.

Tout d'abord, dans cette étude de cas, le but est de prédire si le produit du fournisseur sera en retard ou non. Il est question ici d'un problème de classification binaire. Pour préparer les données, on a supprimé toutes les entrées incomplètes et variables redondantes. Ensuite, toutes les données disponibles ont été converties sous forme numérique et normalisées. Pour normaliser les données, on a utilisé la norme de moindre carré (L2-norm). Alors qu'on dispose de 15 caractéristiques au début, on en obtient en tout 33 suite à l'étape d'ingénierie (*feature engineering*). Pour choisir les meilleurs caractéristiques, on utilise la méthode « cross validation » avec laquelle on obtient le meilleur nombre de caractéristiques qu'est 26. On peut observer la variance du score de prédiction sur la figure suivante :

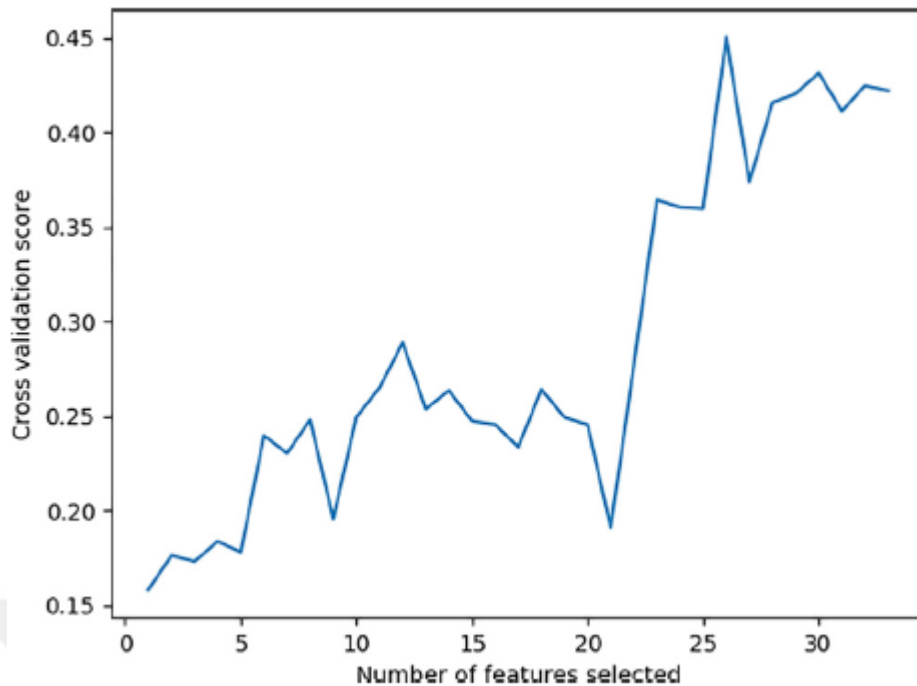


Figure 7: La variance de score de prédiction pour chaque sous-ensemble de 33 caractéristiques.

Pour évaluer les résultats de ce problème, on utilise 4 différentes métriques. Ces métriques sélectionnées conviennent à un ensemble de données déséquilibrées. On sait que :

$$\text{Precision} = \frac{TP}{TP+FP} \text{ et } \text{Recall} = \frac{TP}{TP+FN}$$

Ils sont score de F_1 , précision moyenne, coefficient de corrélation de Matthews et matrice de confusion. On peut trouver les formules de ces métriques en dessous :

1) Score de F_1 :

$$F_1 = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$

2) Précision Moyenne (Average Precision - AP):

$$AP = \sum n(R_n - R_{n-1}) * P_n$$

avec R_n et P_n : indiquant la Precision et Recall au n ème seuil

3) Coefficient de corrélation de Matthews (Matthews Correlation Coefficient-MCC):

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

4) Matrice de Confusion (Confusion Matrix) : afin d'obtenir une vue directe sur la façon dont les chiffres changent entre les vrais/faux positifs et les négatifs (Baryannis, Dani, & Antoniou, 2019).

Pour la sélection des algorithmes, nous devons considérer le trade-off entre la performance et l'interprétabilité. Pour étudier ce trade-off, on choisit d'abord un algorithme très performant pour apprendre « Machines à vecteurs de support ». Deuxièmement, on choisit l'algorithme d'arbre de décision car c'est un algorithme qui peut être facilement interprété même par des non-experts.

Avant d'exécuter l'algorithme, on a divisé notre ensemble de données en deux. On a alloué 20% des données en tant qu'ensemble de test. Et pour déterminer les paramètres optimaux C et γ (paramètre de pénalité et coefficient de kernel), on utilise la recherche par quadrillage (grid search).

On peut trouver un résumé des meilleurs scores de prédiction pour différents classificateurs et métriques en dessous :

Classificateur	AP	F_1	Recall	MCC	Accuracy
Machines à vecteurs de support	0.851	0.791	0.973	0.775	0.943
Arbre de décision	0.698	0.821	0.823	0.792	0.950
Arbre de décision restreint	0.533	0.704	0.928	0.655	0.916

Tableau 5: Un résumé des meilleurs scores de prédiction pour différents classificateurs et métriques

D'après ce tableau, on observe que l'algorithme de machines à vecteurs de support montre la meilleure solution pour les métriques AP et Recall. L'algorithme d'arbre de décision montre la meilleure solution pour les métriques F_1 , MCC et Accuracy. On peut dire que l'arbre de décision sans restriction donne les meilleures solutions par rapport au l'arbre de décision avec restriction.

III. Une approche d'apprentissage machine supervisée pour la simulation pilotée par les données de la sélection de fournisseurs résilients dans la fabrication numérique

Dans cette partie d'étude, nous examinerons l'article suivant : (Cavalcante, Franzon, Forcellini, & Ivanov, 2019)

La définition de la chaîne d'approvisionnement résiliente est :

« La capacité d'adaptation de la chaîne d'approvisionnement pour se préparer aux événements imprévus, réagir aux perturbations et s'en remettre en maintenant la continuité des opérations au niveau souhaité de connectivité et de contrôle de la structure et de la fonction » (Ponomarov & Holcomb, 2009).

Dans cette étude, l'apprentissage machine supervisé est examiné plus en détail en ce qui concerne son application à la sélection des fournisseurs dans la fabrication numérique en tenant compte de la résilience.

Dans un premier temps, nous établirons les profils de risque des fournisseurs et nous sélectionnerons le fournisseur en fonction des profils de risque que nous avons extraits. Nous conceptualisons une approche pour analyser les profils de risque du rendement des fournisseurs en cas d'incertitude en utilisant les capacités d'analyse des données de la fabrication intelligente selon les principes de jumeaux de la chaîne d'approvisionnement numérique.

« Les jumeaux de la chaîne d'approvisionnement numérique ont récemment été définis comme des modèles informatisés qui représentent l'état du réseau à un moment donné et permettent une visibilité complète de la chaîne d'approvisionnement de bout en bout afin d'améliorer la résilience et de tester les plans d'urgence. » (Cavalcante, Franzon, Forcellini, & Ivanov, 2019)

Nous utilisons le modèle de simulation pour les algorithmes d'apprentissage de la source de données à la machine. Dans cette simulation, nous avons quatre fournisseurs. Selon les paramètres du modèle, les commandes de matières premières ne surviennent qu'après la consolidation d'une commande client et la matière première est le seul approvisionnement nécessaire à la fabrication du produit final. Avec les paramètres commande (order) et quantité (quantity) nous choisissons les fournisseurs. Pour cette simulation, on suppose qu'un seul type de produit est livré et on néglige l'analyse des prix et de la concurrence des fournisseurs. La Figure 8 illustre les flux d'informations et de matières dans le modèle de simulation.

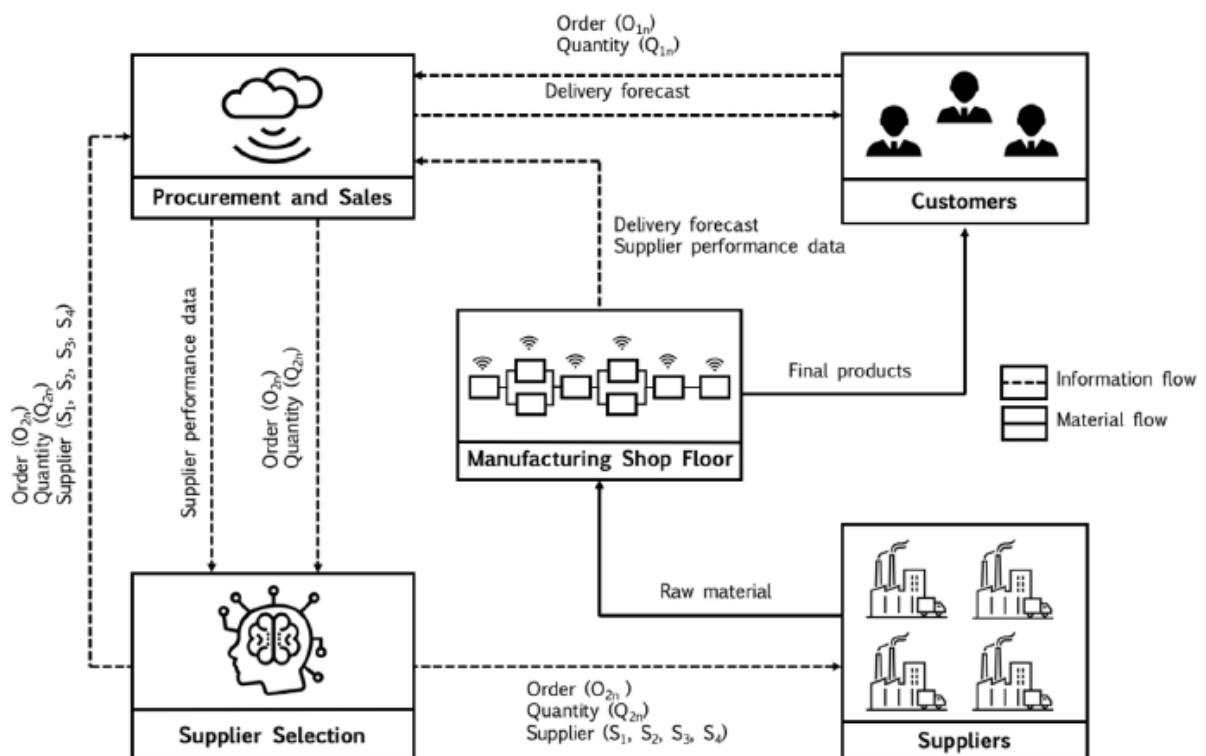


Figure 8: Les flux d'informations et de matières dans le modèle de simulation.

Dans ce travail, le modèle d'apprentissage machine est implémenté en utilisant les bibliothèques Scikit-Learn, Numpy, Matplotlib et Pandas. La Figure 9 montre le modèle de sélection du fournisseur par apprentissage supervisé de la machine.

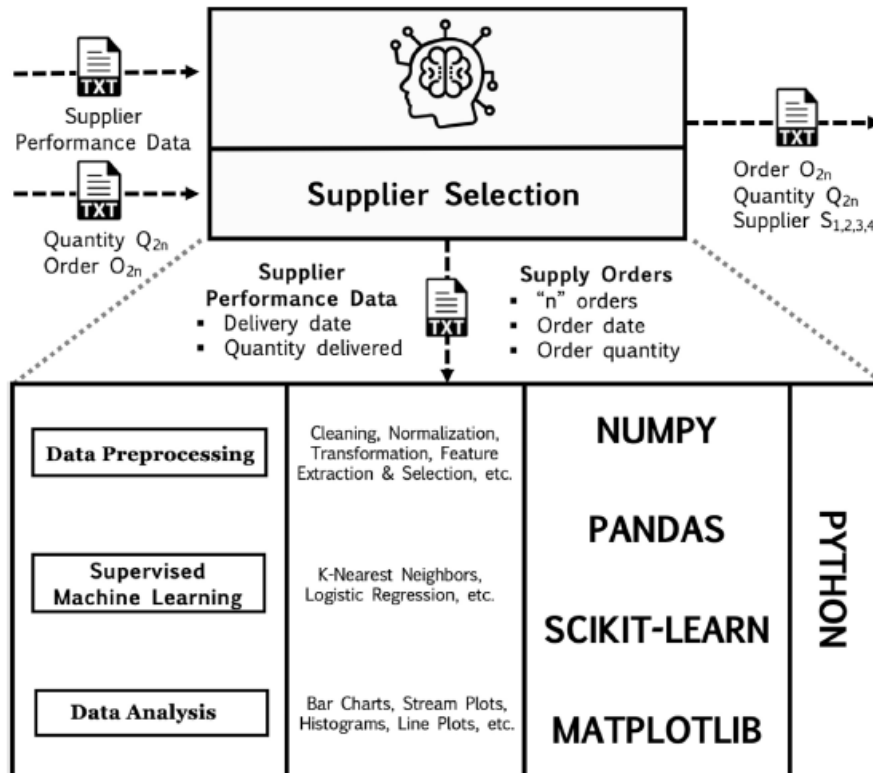


Figure 9: Le modèle de sélection du fournisseur par apprentissage supervisé de la machine

Comme on peut le voir sur la Figure 9, tout d'abord on doit passer par l'étape de prétraitement des données. Durant cette étape, on nettoie, normalise, transforme, extrait et sélectionne les données. Dans ce travail, puisque les données sont générées à partir d'une simulation, la base de données est de bonne qualité : des problèmes tels que des valeurs manquantes, des combinaisons de données impossibles (par exemple, nombre négatif de produits), des valeurs nulles se produisent rarement. Par conséquent, l'étape de prétraitement est plus simple lorsqu'il s'agit des modèles de simulation par rapport aux bases de données réelles. (Cavalcante, Franzon, Forcellini, & Ivanov, 2019)

Ensuite, on utilise deux algorithmes d'apprentissage machine supervisé pour classification des données. Ces derniers sont la méthode des k plus proches voisins et la régression logistique.

Dans ce travail, les données passées sont classées en deux catégories : (i) les livraisons à temps et (ii) les livraisons en retard. L'algorithme k-NN est appliqué séparément pour chacun des deux ensembles de données et mesure la performance du fournisseur selon les caractéristiques du modèle mentionnées précédemment : date de commande et quantité de commande.

L'algorithme de la régression logistique utilise les deux classes de données passées pour les entrées et il donne pour la sortie, la probabilité que chaque fournisseur livre la commande dans le délai prévu pour chaque commande client. L'algorithme de la régression logistique établit un profil de risque pour chaque fournisseur. Ensuite, grâce à un classement des fournisseurs, le fournisseur le moins risqué pour cette commande particulière est sélectionné.

L'algorithme k-NN considère les mêmes données d'entrée et l'algorithme prédit le fournisseur avec la plus grande probabilité de livrer une commande à temps, ainsi que celui avec la plus grande probabilité d'effectuer une livraison tardive.

De plus, une combinaison de ces deux techniques est présentée dans cet article. La première, Hybride A, confronte les résultats de la classification des deux algorithmes sans considérer la précision de chaque technique. La seconde, Hybride B, adopte la même approche, mais tient compte de la précision de chaque classificateur.

On peut trouver les algorithmes de Hybride A et Hybride B ci-dessous :

- acc_a et acc_b représentent la précision du modèle k-NN qui utilise les livraisons à temps et les livraisons en retard, respectivement.
- R_{ka} et R_{kb} représentent les résultats du classificateur k-NN en utilisant respectivement les catégorisations des livraisons à temps et des livraisons en retard.
- R_{lr1} et R_{lr2} représentent les résultats pour le premier et le deuxième fournisseurs les plus susceptibles de répondre à la demande en temps voulu selon le classificateur LR, respectivement.
- La précision pour le modèle de la régression logistique est la surface sous la courbe du caractéristique de fonctionnement de réception (ROC). L'aire sous la courbe (AUC) peut varier de 0 à 1 et une valeur de 0,5 est considérée comme une performance de prédiction aléatoire. (Cavalcante, Franzon, Forcellini, & Ivanov, 2019)

```

1:      procedure SELECTION( $d, q$ ) ▷ date and quantity
2:          if  $R_{lr1} = R_{kb}$  then
3:              return  $R_{ka}$ 
4:          else
5:              return  $R_{lr1}$ 
6:          end if
7:      end procedure

```

Figure 10: L'algorithme de Hybride A.

```

2:      procedure SELECTION( $d, q$ ) ▷ date and quantity
3:          if  $AUC \geq acc_a$  then
4:              if  $AUC \geq acc_b$  then
5:                  return  $R_{lr1}$ 
6:              else
7:                  if  $R_{kb} = R_{lr1}$  then
8:                      return  $R_{lr2}$ 
9:                  else
10:                     return  $R_{lr1}$ 
11:                 end if
12:             end if
13:         else
14:             if  $AUC \geq acc_b$  then
15:                 if  $R_{ka} = R_{kb}$  then
16:                     if  $R_{ka} = R_{lr1}$  then
17:                         return  $R_{lr2}$ 
18:                     else
19:                         return  $R_{lr1}$ 
20:                     end if
21:                 else
22:                     return  $R_{ka}$ 
23:                 end if
24:             else
25:                 return  $R_{ka}$ 
26:             end if
27:         end if
28:     end procedure

```

Figure 11: L'algorithme de Hybride B.

Si on compare la fiabilité des méthodes différentes, d'après le graphe suivant, on observe que Hybride B donne la meilleure solution pour quatre fournisseurs, alors que la logistique régression donne la pire des solutions pour quatre fournisseurs. On peut également observer que l'utilisation mixte des algorithmes de l'apprentissage machine supervisé a conduit à une amélioration de la fiabilité des livraisons des fournisseurs selon la figure suivante.

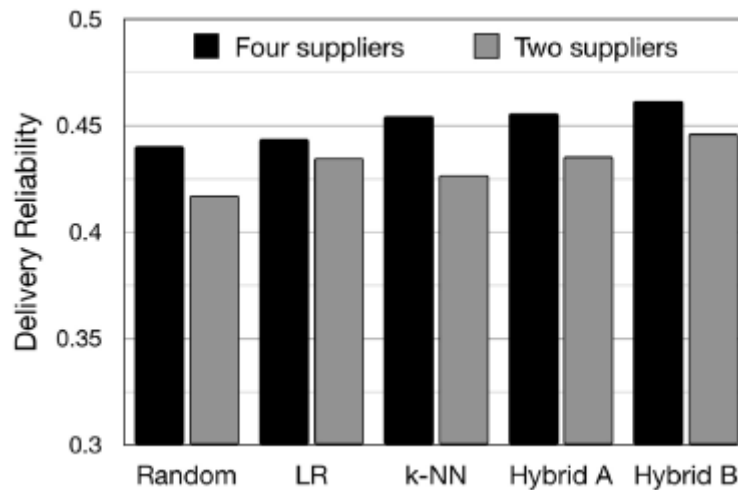


Figure 12: Les fiabilités des méthodes différentes

Conclusion

Pour conclure, ces études nous montrent que l'apprentissage automatique dans la chaîne d'approvisionnement peut être utilisé dans l'évaluation des risques, la prévision de la demande et la sélection des fournisseurs. Ces méthodes peuvent être utilisées pour soutenir de nombreuses décisions dans la gestion de la chaîne d'approvisionnement. De meilleurs résultats peuvent être obtenus en ayant recours à des méthodes d'apprentissage automatique avec d'autres technologies.

Bibliographie

- Baryannis, G., Dani, S., & Antoniou, G. (2019). Predicting supply chain risks using machine learning: The trade-off between performance and interpretability. *Future Generation Computer Systems, 101*, 993–1004.
- Baryannis, G., Validi, S., Dani, S., & Antoniou, G. (2018). Supply chain risk management and artificial intelligence: state of the art and future research directions. *International Journal of Production Research, 57*(7), 2179-2202.
- Business Dictionary*. (2019). Récupéré sur <http://www.businessdictionary.com/definition/na-ve-forecasting.html>
- Carbonneau, R., Laframboise, K., & Vahidov, R. (2008). Application of machine learning techniques for supply chain demand forecasting. *European Journal of Operational Research, 184*, 1140–1154.
- Cavalcante, I. M., Franzon, E. M., Forcellini, F. A., & Ivanov, D. (2019). A supervised machine learning approach to data-driven simulation of resilient supplier selection in digital manufacturing. *International Journal of Information Management, 49*, 86-97.
- Chopra, S., & Meindl, P. (2001). *Supply Chain Management*. Prentice Hall.
- Christopher, M. (1994). *Logistics and Supply Chain Management*. New York: Pitman Publishing.
- CSCMP Supply Chain Management Definitions and Glossary*. (2013). Récupéré sur https://cscmp.org/CSCMP/Educate/SCM_Definitions_and_Glossary_of_Terms.aspx
- Hayes, A. (2019). *Investopedia*. Récupéré sur <https://www.investopedia.com/terms/m/movingaverage.asp>
- Kenton, W. (2019). *Investopedia*. Récupéré sur <https://www.investopedia.com/terms/m/mlr.asp>
- Kullback, S. (1968). *Information Theory and Statistics*. New York: Dover Books.
- Ponomarov, S. Y., & Holcomb, M. C. (2009). Understanding the concept of supply chain resilience. *International Journal of Logistics Management, 20*(1), 124-143.
- Sila, I., Ebrahimpour, M., & Birkholz, C. (2006). Quality in supply chains: an empirical analysis. *Supply Chain Management, 11*(6), 491-502.
- Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. New York: Springer.
- Werbos, P. J. (1990). Backpropagation through time: What it does. *Proceedings of the IEEE, 78*(10), 1550 - 1560.