

# Méthodologie de prévision des ventes B2B avec une demande intermittente

HAITHAM MZOUGHJI<sup>1</sup>, MAHA BENALI<sup>1</sup>, JEAN-MARC FRAYRET<sup>1</sup>

<sup>1</sup> DEPARTEMENT DE MATHÉMATIQUES ET DE GENIE INDUSTRIEL

POLYTECHNIQUE MONTREAL, CANADA

[haithem.mzoughi@polymtl.ca](mailto:haithem.mzoughi@polymtl.ca) [maha.benali@polymtl.ca](mailto:maha.benali@polymtl.ca) [jean-marc.frayret@polymtl.ca](mailto:jean-marc.frayret@polymtl.ca)

---

Dans le domaine de la prévision des ventes, le phénomène de demande intermittente pose un défi important pour la précision de la prévision. Une demande intermittente est sporadique et irrégulière. Les méthodes traditionnelles telles que les moyennes mobiles ou le lissage exponentiel ne sont pas capables de les prévoir. Le manque de données historiques et l'influence de facteurs externes peuvent rendre ces prévisions d'autant plus difficiles à faire. Pour faire face à ces défis, diverses méthodes avancées de prévision ont été proposées dans la littérature. Parmi ces méthodes, la méthode de Croston a été spécifiquement développée pour prévoir la demande intermittente. Les méthodes bayésiennes sont aussi connues à être efficaces pour modéliser l'incertitude et l'irrégularité de la demande. L'incorporation de données externes telles que les tendances de l'industrie ou du marché pourrait également améliorer la précision de la prévision.

Cette étude vise à contribuer à cette littérature en étudiant à l'aide d'un cas industriel l'efficacité de ces différentes méthodes avancées à prévoir les ventes B2B pour les produits avec une demande intermittente. En outre, cette étude examine l'impact de l'incorporation de données externes sur la précision de la prévision. Nous proposons une application innovante pour améliorer la prévision en commençant par prédire quand la vente aura lieu. À cette fin, nous utilisons deux approches, soit une basée sur des modèles de classification et une basée sur l'analyse de séries temporelles des intervalles interdemande. Nous proposons également d'utiliser une approche d'agrégation/désagrégation temporelle afin de fournir une prévision plus précise de la demande. Les résultats fournissent des informations pertinentes pour les praticiens et les chercheurs dans le domaine de la prévision des ventes B2B, et peuvent contribuer au développement de stratégies de prévision plus robustes pour la demande intermittente.

**Mots clés** - Prévision des ventes B2B, Demande intermittente, Méthode de Croston, Intervalles interdemandes, Agrégation et désagrégation temporelle.

---

## 1 INTRODUCTION

La prévision de la demande dans un contexte B2B est différente de celle dans un contexte B2C. Les différences incluent des cycles de demande plus longs et complexes, des volumes de ventes plus élevés, et une grande variété de données disponibles. Les méthodes de prévision variées, telles que les approches statistiques, les modèles autorégressifs, les modèles de lissage et les modèles d'apprentissage automatique, peuvent être utilisées pour prendre en compte les facteurs tels que l'agrégation temporelle, l'agrégation des produits et d'autres facteurs d'agrégation de la chaîne d'approvisionnement. L'intermittence, qui est la présence de valeurs nulles dans la série temporelle de la demande, peut également impacter le choix de la méthode de prévision.

Dans cette étude, nous nous intéressons spécifiquement au cas d'un distributeur disposant de 6 entrepôts localisés dans la province du Québec. Le distributeur offre plus de 50 000 produits de boulonnerie et de quincaillerie industrielle, vendus principalement à des clients industriels localisés dans les provinces de Québec et d'Ontario, ainsi que dans l'Est des États-Unis. D'après l'historique de la demande, on remarque que certains produits sont commandés régulièrement, tandis que pour la grande majorité, ils sont commandés moins d'une fois par mois. Ceci représente un enjeu pour les dirigeants qui souhaiteraient prévoir avec précision les niveaux de ventes mensuelles dans le but d'optimiser les activités de réapprovisionnement des stocks.

Cette étude contribue à la littérature existante sur la prévision de la demande intermittente en étudiant à l'aide d'un cas industriel

l'efficacité de différentes méthodes avancées à prévoir les ventes des produits à demande intermittente à partir de l'historique des ventes du partenaire industriel. Dans ce cas d'étude, l'expression « prévision de la demande » est considérée comme l'équivalente de « prévision des ventes » vu que le distributeur arrive généralement à combler toute la demande reçue.

L'objectif de l'étude est de développer une méthodologie de prévision des ventes mensuelles dans un contexte B2B pour les produits à demande intermittente. Cet article est structuré comme suit : à la suite de l'introduction, Section 2 présente les méthodes utilisées dans la littérature pour prévoir la demande intermittente. Section 3 décrit la méthodologie que nous avons adoptée pour prévoir la demande intermittente. Dans Section 4, la phase d'exploration et de traitement des données afin de les rendre adaptées pour la modélisation est présentée. Section 5 est consacrée à la présentation des résultats et à la comparaison de la performance des différentes approches de prévision testées. Une conclusion et les recommandations au partenaire industriel sont présentées dans Section 6.

## 2 REVUE DE LITTÉRATURE

Nous pouvons classer les méthodes de prévision de la demande intermittente en trois catégories : les méthodes basées sur le modèle de Croston, les méthodes basées sur l'agrégation, et les méthodes utilisant l'apprentissage automatique.

### 2.1 Modèle Croston

La méthode de Croston (1972) utilise des séries temporelles intermittentes. Elle utilise un lissage exponentiel pour estimer la demande moyenne future (Willemain et al. 1994). La méthode d'Approximation Syntetos-Boylan (SBA) est une extension de

la méthode de Croston qui introduit un deuxième paramètre pour prendre en compte la probabilité de demande nulle (Syntetos, Boylan, and Croston 2005). Cette méthode repose sur deux hypothèses : premièrement, que la demande est intermittente et deuxièmement, que les intervalles entre les périodes de demande non nulle suivent une distribution binomiale négative. La méthode SBA estime le volume de demande attendu et l'intervalle attendu entre les demandes de manière distincte, puis combine ces estimations pour produire une prévision pour la période suivante (Syntetos, Boylan, and Croston 2005). Une autre amélioration du modèle de Croston a été introduite afin de permettre au modèle de prendre en compte la mise à jour des paramètres après plusieurs périodes de demande nulle (Teunter, Syntetos, and Babai 2011). Plus récemment, Feng et al. (2021) ont proposé une autre modification de la méthode de Croston en utilisant un algorithme génétique pour optimiser les paramètres, ce qui a abouti à des prévisions plus précises.

Malgré sa popularité, le modèle de Croston présente des limites en ce qui concerne la prévision de données de demande intermittente. Premièrement, il suppose des probabilités constantes pour l'occurrence de la demande et les intervalles entre les demandes, ce qui n'est pas toujours le cas dans la pratique. Deuxièmement, ce modèle suppose une indépendance entre les tailles de demande et les intervalles entre les demandes, ce qui peut ne pas être vrai dans certains cas. Troisièmement, ce modèle ne prend pas en compte la distribution des tailles de demande. Enfin, le modèle de Croston peut ne pas fonctionner bien pour un volume de données insuffisant pour estimer avec précision les probabilités d'occurrence de la demande et des intervalles entre les demandes (Xu, Wang, and Shi 2012).

## 2.2 Prévision par agrégation

La prévision par agrégation pourrait être réalisée de deux façons (Spiliotis et al. 2020). La première est la consolidation d'indices de temps, comme consolider la demande hebdomadaire pour construire une série temporelle mensuelle de demande, ou consolider la demande quotidienne pour construire la demande hebdomadaire. La deuxième façon de faire est de réaliser une agrégation hiérarchique en regroupant la demande des produits appartenant à la même catégorie, comme la demande livrée au même client par le même entrepôt.

### 2.2.1 Prévision par agrégation temporelle

L'agrégation temporelle consiste à regrouper des unités de temps inférieures (comme les jours) pour construire des unités de temps supérieures (comme les semaines ou les mois). Cela permet de réduire l'effet de la discontinuité dans les données intermittentes. Spithourakis et al. (2014) ont proposé une approche d'agrégation et de désagrégation qui combine les techniques d'agrégation avec des techniques de prévision pour une meilleure précision. Cette approche a été utilisée par Petropoulos and Kourentzes (2015) dans le cadre du modèle Croston, ainsi que par Lei, Li, and Tan (2016) qui utilisent les chaînes de Markov floue pour estimer les coefficients de désagrégation.

La méthode d'estimation de la demande intermittente par agrégation temporelle performe moins dans le cas de demande à forte variabilité. De plus, elle est moins adaptée à la prévision à court terme, car l'exactitude de la prévision pourrait être affectée par la durée de l'intervalle d'agrégation. Enfin, cette méthode ne prend pas en compte l'impact des facteurs externes sur la demande (ex., les promotions, les conditions du marché (Babai, Ali, and Nikolopoulos 2012).

### 2.2.2 Prévision par agrégation hiérarchique

L'agrégation hiérarchique permet de prévoir la demande en agrégeant les données à différents niveaux de la chaîne d'approvisionnement (p.ex., agréger les produits en catégories de produits, ou encore agréger les données définies par magasin en des données par ville ou pays). Les prévisions peuvent être générées soit par la méthode de bas en haut, en commençant par faire des prévisions pour les niveaux inférieurs, puis en les sommant pour obtenir des prévisions pour les niveaux supérieurs, soit par la méthode de haut en bas, en commençant par faire des prévisions pour un niveau supérieur et en déduisant les prévisions pour les niveaux inférieurs à l'aide de coefficients de désagrégation. Les deux méthodes peuvent donner des prévisions raisonnables au niveau inférieur (Syntetos et al. 2016).

Athanasopoulos, Ahmed, and Hyndman (2009) ont proposé un modèle pour décomposer les prévisions du haut niveau en fonction des proportions prévues pour les séries de temps inférieures, plutôt que des proportions historiques. Abolghasemi et al. (2019) ont proposé une approche de prévision hiérarchique qui utilise les modèles Xgboost, SVM et réseaux de neurones, et qui performe mieux que le modèle ARIMAX en prenant le prix en tant que variable exogène. L'approche proposée prédit la demande des vendeurs d'un niveau moyen dans la chaîne d'approvisionnement en utilisant les séries temporelles, des coefficients d'agrégation du niveau inférieur, ainsi que des coefficients de désagrégation du niveau supérieur. Cependant, la précision de la prévision peut être altérée par le choix de la hiérarchie de regroupement et les modèles utilisés pour chaque niveau peuvent ne pas être adaptés à tous les produits. En outre, l'approche peut être très gourmande en termes de calculs et peut ne pas être sensible aux changements brusques dans les modèles de demande ou les fluctuations des conditions du marché. (Kourentzes and Athanasopoulos 2021).

### 2.3 Prévision de la demande intermittente par apprentissage automatique

Kourentzes (2013) a proposé une structure spécifique de réseau de neurones pour la prévision de la demande intermittente. Il a utilisé la corrélation entre la série de demandes non-nulles et celle des intervalles interdemandes (durées entre 2 demandes non-nulles) pour prédire le volume de la prochaine demande et le prochain intervalle interdemande. La structure du réseau avait une couche cachée avec une fonction d'activation tangente hyperbolique et une couche de sortie avec deux neurones ayant une fonction d'activation linéaire.

Les approches pour la prévision de la demande intermittente ont évolué au fil du temps avec des propositions différentes. Nikolopoulos, Babai, and Bozos (2016) ont développé un cadre conceptuel basé sur l'approche des plus proches voisins en utilisant la distance euclidienne pour prévoir les demandes intermittentes. Turkmen, Wang, and Januschowski (2019) ont supposé que les intervalles interdemandes et les volumes suivent une distribution binomiale négative et ont utilisé un réseau de neurones pour estimer les paramètres de cette distribution. Liu et al. (2020) ont proposé une structure de deux modèles LSTM pour prédire à la fois l'intervalle et le volume.

### 2.4 Méthodes proposées

Cette étude propose d'utiliser des approches basées sur l'apprentissage automatique pour prévoir la demande intermittente. La première approche consiste à prédire la présence ou l'absence d'une demande en utilisant XGboost, un algorithme connu par sa précision et par sa robustesse vis-à-vis

aux valeurs aberrantes (Wang, Deng, and Wang 2020), et SVM (Support Vector Machine), un algorithme de classification connu par son aptitude à séparer les données qui ne sont pas séparables linéairement et sa robustesse vis-à-vis aux valeurs aberrantes (Awad et al. 2015). La deuxième approche, inspirée de Skariah and Suriyakala (2022), utilise les séries des intervalles interdemandes pour prévoir quand une demande aura lieu. Inspirée de Gourieroux and Jasiak (2004), la troisième approche procède par agrégation/désagrégation en utilisant un coefficient de désagrégation prédit.

### 3 METHODOLOGIE DE RECHERCHE

Cette étude a pour objectif de développer une méthodologie de prévision des ventes mensuelles dans un contexte B2B pour les produits à demande intermittente. La méthodologie est structurée autour des étapes présentées dans Figure 1.

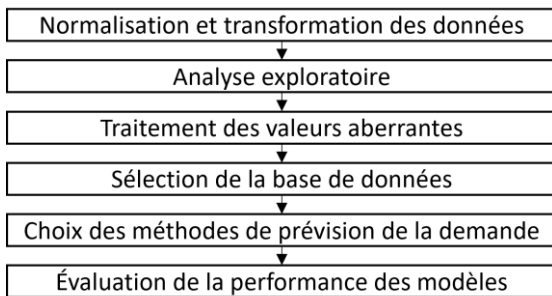


Figure 1 : Méthodologie de recherche

#### 3.1 Normalisation et transformation des données

La normalisation et la transformation des données consistent à extraire les données transactionnelles utiles (ID clients, ID produits, quantités, dates, prix unitaires), à supprimer les redondances, à traiter les valeurs manquantes et à créer une base de données standardisée pour la modélisation. Notre source de données est le système ERP de l'entreprise.

#### 3.2 Analyse exploratoire

L'analyse exploratoire vise à extraire des informations pertinentes à partir des données brutes pour évaluer la qualité des données et répondre à des questions clés telles que les produits/catégories les plus vendus et les plus rentables, les produits à demande intermittente, ainsi que la distribution de la fréquence des ventes.

#### 3.3 Traitement des valeurs aberrantes

Les valeurs aberrantes peuvent être présentes dans les données en raison de la possibilité d'erreur de saisie dans les champs libres du système ERP. Pour garantir la précision de la prévision, il est important de traiter les valeurs aberrantes séparément lors de la normalisation des données. Nous traitons chaque série temporelle de demande de produit de façon individuelle en utilisant des méthodes de détection appropriées. Les méthodes de détection des valeurs aberrantes utilisées sont :

- 1- L'InterQuartile Range : l'IQR mesure la variabilité d'une série de données. Les valeurs en dehors de l'intervalle défini par  $Q1 - 1,5 * IQR$  et  $Q3 + 1,5 * IQR$  sont considérées comme aberrantes,  $Q1$  et  $Q3$  étant respectivement les quartiles inférieur et supérieur.
- 2- La méthode Density-Based Spatial Clustering of Applications with Noise : DBSCAN est un algorithme de clustering de données qui peut être utilisé pour détecter les valeurs aberrantes dans les séries temporelles. Il se base sur la densité des données pour regrouper les points qui sont proches les uns des autres en un cluster et d'identifier les points éloignés (Tripathy and Sahoo 2020).
- 3- La méthode Local Outlier Factor : la LOF est une méthode

de détection d'anomalies utilisée dans les séries temporelles. Le facteur LOF associé à une observation  $i$  et calculé par l'équation (1) mesure la similarité relative entre  $i$  et ses voisins les plus proches.

$$LOF(i) = \frac{1}{\frac{1 - \text{densité relative}(i)}{\text{moyenne des densités relatives des voisins de } i}} \quad (1)$$

Les voisins ( $i$ ) sont les  $k$  observations les plus proches de  $i$ . La densité relative ( $i$ ) mesure la densité de voisinage autour de  $i$  par rapport à la densité totale des données. La moyenne des densités relatives des voisins de  $i$  mesure la moyenne de la densité de voisinage autour des  $k$  voisins les plus proches de  $i$ . Si LOF est supérieur à 1, le point est jugé aberrant.

#### 3.4 Sélection de la base de données

Nous étudions se limite aux produits qui contribuent les plus au chiffre d'affaires de tout l'historique de l'entreprise. Ainsi, nous avons utilisé la méthode ABC afin de classer les produits en fonction de leur contribution au chiffre d'affaires total. Ceci a permis d'identifier les produits de la Classe A (générant 80% du chiffre d'affaires) et de se limiter à l'analyse de 3717 produits parmi 53 284 produits actifs (c-à-d., vendus entre 2019 et 2020).

Puisque certains produits peuvent être vendus en vrac ou en boîte, la variable quantité vendue dans la base de données peut soit refléter le nombre des boîtes, soit le nombre d'unités de produits. Pour éviter toute confusion, nous avons décidé d'utiliser les séries temporelles des ventes mensuelles en termes de chiffres d'affaires associées aux ventes. Dans ce qui suit, toutes les méthodes sont déployées afin de prévoir le chiffre d'affaires généré par les ventes prévues (les ventes étant équivalentes à la demande tel qu'est expliqué dans Section 1).

#### 3.5 Choix des méthodes de prévision des ventes

Puisque nous ne possédons pas d'outil qui nous permette de juger si une demande d'un produit est intermittente, nous avons testé d'abord des méthodes de prévision de référence (ne tiennent pas compte du phénomène d'intermittence), puis des approches dédiées à la prévision de la demande intermittente.

##### 3.5.1 Méthodes de prévision de référence

Nous avons testé des *méthodes de prévision de référence* comme la méthode ARIMA et la méthode de Holt-Winters, ainsi que la méthode basée sur le *modèle de Croston*.

##### 3.5.2 Méthodes hybrides pour la prévision de la demande intermittente

Nous avons testé *trois approches hybrides* pour affiner la prévision de la demande intermittente du partenaire. La *première approche hybride* testée (voir Figure 2) est composée de deux étapes : (1) la prédiction de l'occurrence d'une demande, soit un problème de classification binaire (1 si une demande aura lieu, 0 sinon) ; et (2) la prévision du chiffre d'affaires (\$) généré par les ventes en utilisant les séries temporelles des ventes non nulles. À l'étape 1, à partir des séries temporelles des ventes, nous avons créé des variables permettant d'entraîner des modèles de classification binaire pour prédire la présence de ventes dans une unité de temps (un mois dans notre cas). Nous avons utilisé des variables telles que la présence d'une demande à la semaine précédente, la présence d'une demande à la même semaine mais de l'année précédente, la présence d'une demande au mois courant, le total des ventes depuis le début de l'année jusqu'à la semaine en question, et les ventes totales de l'année précédente. Les deux modèles de classification XGBoost et SVM ont été utilisés à cette étape. À l'étape 2, pour prévoir le chiffre d'affaires généré par les ventes, nous avons utilisé la série des ventes non-nulles pour entraîner

des modèles de prévision tels que ARIMA, SVR et LSTM. Finalement, la prévision du chiffre d'affaires est calculée comme étant le produit des résultats des deux modèles, soit le résultat de la classification binaire multiplié par le chiffre d'affaires prévu par la série des ventes non nulles.

La *deuxième approche hybride* utilisée (voir Figure 3) se fait également en deux étapes. L'étape 1 vise à prédire les intervalles interdemandes (correspondant aux intervalles entre les ventes dans notre cas). Ainsi, au lieu d'utiliser des modèles de classification comme dans l'approche précédente, nous avons utilisé des modèles de prévision de séries temporelles tels que ARIMA, le lissage exponentiel et LSTM. Dans l'étape 2, les séries temporelles des ventes non nulles ont été utilisées pour prédire le chiffre d'affaires comme pour l'approche précédente.

Pour résumer, l'étape 1 pour les deux premières approches hybrides vise à prévoir quand la demande aura lieu, mais utilise des outils différents (la classification pour la première approche et la prévision des intervalles interdemandes pour la deuxième).

Pour pouvoir comparer les résultats de l'étape 1 pour ces deux approches, nous avons transformé les résultats des modèles de prévision des intervalles interdemandes en vecteurs binaires (1 si une demande aura lieu, 0 sinon). Ceci permettra d'utiliser les mêmes métriques de précision (F1-score et l'exactitude).

Finalement, la *troisième approche hybride* (voir Figure 4) tente de réduire la rareté des séries temporelles des ventes en procédant par agrégation/désagrégation. Le chiffre d'affaires généré par les ventes agrégées est alors prédit, puis désagrégé à l'aide d'un coefficient de désagrégation estimé par un modèle INAR. Le choix de ce modèle est inspiré de Nikolopoulos et al. (2011). Comme le modèle ARIMA, le modèle INAR est utilisé pour la prévision de séries temporelles, mais il est spécialement conçu pour prévoir des séries temporelles comprenant des dénombrements ou des observations à valeurs entières. L'estimation des coefficients d'un modèle INAR se fait à l'aide de la méthode du maximum de vraisemblance ou de méthodes bayésiennes (Alwan and Weiß 2017).

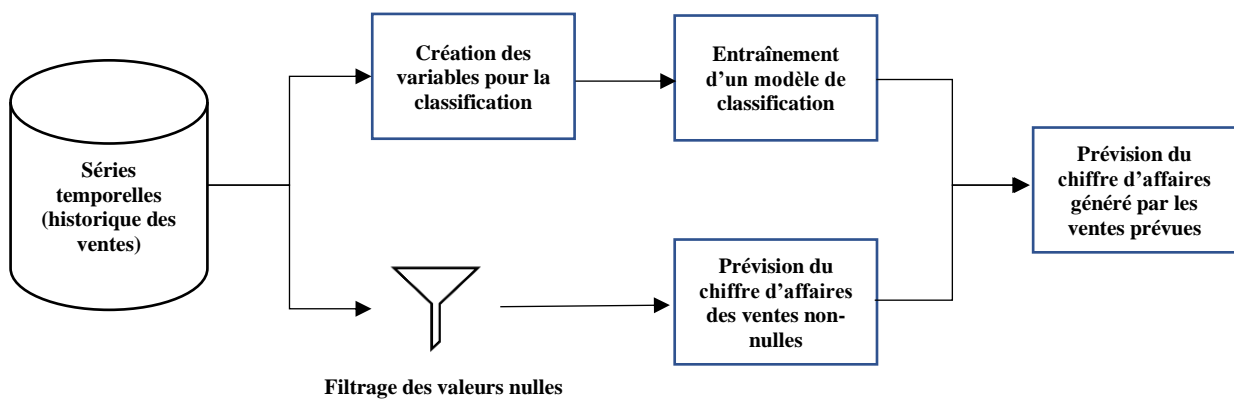


Figure 2 : Processus de prévision par classification

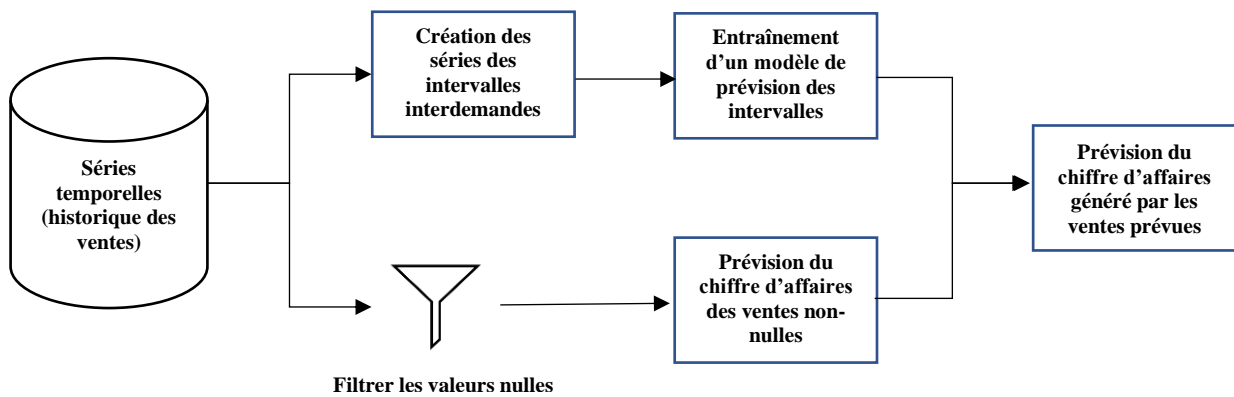


Figure 3 : Processus de prévision par intervalles interdemandes

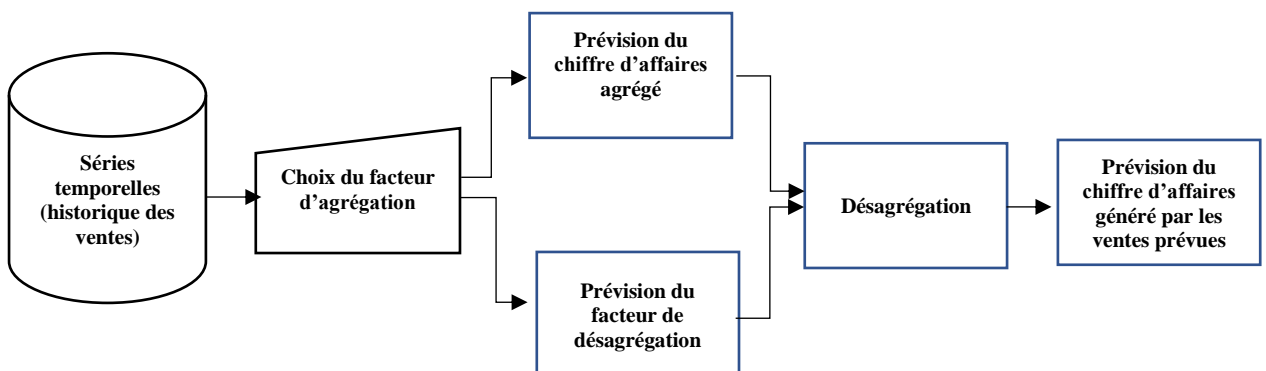


Figure 4 : Processus de prévision par agrégation et désagrégation

### 3.6 Évaluation de la performance des modèles

Pour évaluer la performance des différents modèles décrits précédemment, nous avons utilisé les données transactionnelles du partenaire de 2007 jusqu'à la fin de 2019. Plus particulièrement, nous avons utilisé les données de 2019 comme données de validation, et les données de 2007 à 2018 pour entraîner les modèles. Les indicateurs utilisés sont les suivants :

- Le MAPE (Mean Absolute Percentage Error) mesure la moyenne des erreurs absolues en pourcentage et est calculé par l'équation (2).

$$MAPE = \frac{1}{n} \left( \frac{\sum |y_i - \hat{y}_i|}{y_i} \right) \quad (2)$$

où  $n$  est le nombre d'observations,  $y_i$  est la valeur réelle de la série temporelle, et  $\hat{y}_i$  est la valeur prédite (Chen, Twycross, and Garibaldi 2017).

- Le sMAPE (Symmetric Mean Absolute Percentage Error) est une variante du MAPE qui permet d'éviter le problème de la division par zéro (Chicco, Warrens, and Jurman 2021) et qui est calculée par l'équation (3).

$$sMAPE = \frac{2}{n} \left( \frac{\sum |y_i - \hat{y}_i|}{y_i + \hat{y}_i} \right) \quad (3)$$

Comme pour certains produits, les données du test contiennent des valeurs nulles, nous avons dû utiliser le sMAPE pour les raisons suivantes :

- 1- L'interprétation est plus claire : le sMAPE a une gamme de valeurs de 0% à 200%, ce qui est plus facile à interpréter que les valeurs infinies potentielles du MAPE.
- 2- Les valeurs sont absolues : le sMAPE utilise des valeurs absolues pour le calcul de l'erreur, ce qui signifie qu'il n'y a pas de biais en faveur des erreurs de sous-estimation ou de surestimation.
- 3- Les valeurs nulles sont bien gérées : le MAPE peut donner des résultats indéfinis ou non pertinents lorsque la valeur observée est égale à zéro. Cependant, le sMAPE a une approche différente pour gérer les valeurs nulles, ce qui le rend plus fiable pour les séries temporelles qui contiennent des valeurs nulles. (Chen, Twycross, and Garibaldi 2017).

Pour les approches qui incluent une étape de classification, nous avons considéré quatre indicateurs dédiés à l'évaluation de la performance des modèles de classification.

- La Précision : elle est calculée par l'équation (4) et mesure la proportion des prédictions positives (Yacouby and Axman 2020).

$$\text{Précision} = \frac{\text{Vrais positifs}}{\text{Vrais positifs} + \text{Faux positifs}} \quad (4)$$

- Le Rappel : il est calculé par l'équation (5) et mesure la capacité du classificateur à détecter toutes les instances positives (Yacouby and Axman 2020):

$$\text{Rappel} = \frac{\text{Vrais positifs}}{\text{Vrais positifs} + \text{Faux négatifs}} \quad (5)$$

- L'exactitude : il s'agit d'une métrique couramment utilisée pour évaluer les performances des classificateurs binaires. Telle que présentée par l'équation (6), elle mesure la proportion d'instances qui sont correctement classées.

$$\text{Exactitude} = \frac{\text{Vrais positifs} + \text{Vrais négatifs}}{\text{Nombre total d'instances}} \quad (6)$$

Le terme « Vrais positifs » désigne le nombre d'instances qui sont réellement positives et sont correctement prédites comme positives par le classificateur. Le terme « Faux positifs » désigne le nombre d'instances qui sont réellement négatives, mais qui sont prédites comme positives par le classificateur. Le terme Faux négatifs désigne le nombre d'instances qui sont réellement positives, mais qui sont prédites comme négatives par le classificateur (Novaković et al. 2017).

- Score F1 (Tan et al. 2019) : calculée par l'équation (7), cette métrique est utilisée pour évaluer la performance d'un classificateur binaire. Elle assure un compromis entre la précision et le rappel.

$$\text{Score F1} = 2 \frac{\text{Précision} * \text{Rappel}}{\text{Précision} + \text{Rappel}} \quad (7)$$

## 4 ANALYSE EXPLORATOIRE ET PRETRAITEMENT DES DONNEES

Comme cette étude de cas analyse des données réelles d'un partenaire industriel, beaucoup de temps a été consacré au prétraitement des données qui est crucial pour la qualité de la prévision, quelle que soit la méthode utilisée. Le jeu de données utilisé contient des données sur les produits, des données sur les clients et des données des commandes. Les données sur les produits comprennent un identifiant unique pour chaque produit, son nom, sa description, et sa catégorie. Les données des commandes comprennent la quantité commandée, le prix unitaire, le total, ainsi que les clés externes pour le produit et le client associés à la commande.

### 4.1 Construction des séries temporelles

L'objectif de cette étape est d'identifier les produits pour lesquels nous pouvons construire des séries temporelles pour la prévision en fonction de leur occurrence des ventes. L'occurrence des ventes désigne le nombre de fois qu'un produit a été vendu dans l'historique des ventes analysé. Après l'élimination des produits non-codés (c-à-d., les produits qui n'ont pas d'identifiants dans la base de données), 72% des produits restants ont une occurrence d'achat inférieure à 10 fois, avec la plupart d'entre eux ayant une fréquence d'achat égale à 1. Plus de 34 900 produits ont été vendus une seule fois depuis l'installation de l'ERP, seulement 36 000 produits sont vendus plus de 10 fois. De plus, plus de 75 000 produits ont été vendus pour la dernière fois en 2018 ou avant. Nous rappelons que nous souhaitons prévoir les ventes de 2019 (ensemble pour la validation).

### 4.2 Traitement des valeurs aberrantes

Les données de l'ERP sur les prix et les quantités sont saisies manuellement. Comme les erreurs de saisie sont fréquentes, le traitement des valeurs aberrantes est crucial. Pour les détecter, nous avons eu recours à trois approches différentes présentées dans la sous-section 3.3. Ensuite, nous avons appliqué un modèle de lissage exponentiel Holt-Winters, qui est sensible aux valeurs aberrantes dans les données. Les erreurs absolues moyennes obtenues avec les trois approches de traitement des valeurs aberrantes sont synthétisées par Figure 5. La méthode LOF a donné l'erreur absolue moyenne la plus faible. Par conséquent, nous l'avons utilisée pour le reste de notre étude.

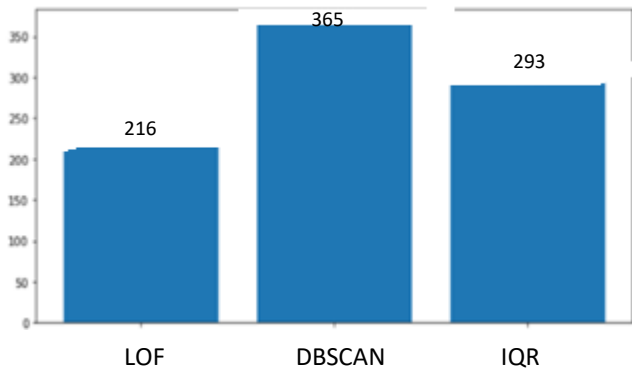


Figure 5 : Erreurs absolues moyennes obtenues avec les trois approches de traitement des valeurs aberrantes

### 4.3 Sélection des produits

Nous avons effectué une sélection rigoureuse des données sources en utilisant différents filtres. Sur un ensemble de 129193 produits dans la base de données, nous avons gardé seulement 53 284 produits actifs (c-à-d., vendus entre 2019 et 2020). Ensuite, nous avons exclu les produits atypiques liés à la période de la COVID-19. Enfin, nous avons utilisé la méthode LOF pour éliminer les valeurs aberrantes sur les séries temporelles de ventes en termes de chiffre d'affaires. Finalement, nous avons utilisé la méthode ABC pour se limiter à l'analyse de 3 717 produits représentant 80% du chiffre d'affaires total (entre 2007 et 2019) du partenaire.

## 5 ANALYSE DES RESULTATS

Nous commençons par tester les méthodes de référence et analysons l'effet de l'intermittence sur l'erreur de prévision obtenue avec ces méthodes pour justifier le recours aux méthodes hybrides. Ensuite, nous présentons une comparaison de la performance des trois méthodes hybrides proposées dans Section 3. L'idée est de déterminer quelle approche de prévision des ventes donne les meilleurs résultats pour chaque produit. Dans ce qui suit, seuls les 3717 produits de la classe A ont été considérés.

### 5.1 Effet de l'intermittence sur l'erreur de prévision avec les méthodes de référence

Nous avons testé la performance des trois méthodes de référence (ARIMA, Holt-Winters et Croston) pour chaque produit. Figure 6 illustre l'effet du taux de rareté (et donc de l'intermittence) sur l'erreur de prévision pour ces méthodes. Par exemple, le point bleu sur la figure représente la sMAPE moyenne sur tous les produits ayant un taux de rareté de 0.2 obtenue avec le modèle ARIMA. Figure 6 montre que l'erreur de prévision est davantage plus importante pour les produits ayant historiquement réalisé moins de ventes (donc pour lesquels la demande est fortement intermittente).

Figure 7 montre que les trois méthodes sont similaires en termes de performance prédictive, étant donné qu'il y a un chevauchement important entre les boîtes à moustaches. L'erreur de prévision moyenne pour les trois méthodes est élevée (de l'ordre de 60 %), ce qui nous incite à tester les trois approches hybrides (voir Section 3.5.2).

### 5.1 Comparaison de l'approche par classification et l'approche par intervalles interdemandes

Pour ces deux approches, la prévision se fait en deux étapes, tel que décrit dans Section 3.5.2.

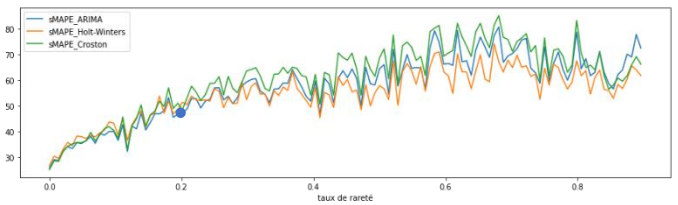


Figure 6 : Effet de l'intermittence sur l'erreur de prévision (sMAPE) obtenue avec les méthodes (ARIMA, Holt-Winters et Croston)

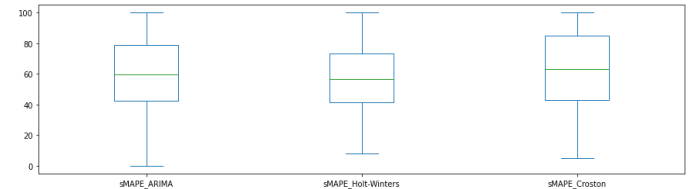


Figure 7 : Comparaison des diagrammes en boîte à moustaches de l'erreur MPAE moyenne sur tous les produits obtenus avec les méthodes (ARIMA, Holt-Winters et Croston)

### 5.1.1 Comparaison de la performance à prévoir l'occurrence des ventes (Étape 1)

Nous proposons de comparer la performance de ces deux approches à l'Étape 1 (soit à prévoir quand la vente aura lieu, ou ce qu'on appelle l'occurrence des ventes) vu qu'elles utilisent des outils différents (la classification pour la 1<sup>ère</sup> approche et la prévision des intervalles interdemandes pour la 2<sup>ème</sup>). Pour pouvoir comparer les résultats de l'Étape 1 et utiliser les mêmes métriques (F1-score et l'exactitude), nous avons transformé les résultats des modèles de prévision des intervalles interdemandes en vecteurs binaires (1 si une vente aura lieu, 0 sinon).

Figures 8 et 9 présentent respectivement le F1-score et l'exactitude des prévisions de l'occurrence des ventes obtenues avec l'approche par classification utilisant les algorithmes XGBoost et SVM, ainsi qu'avec l'approche par intervalles interdemandes utilisant les algorithmes ARIMA, Holt-Winters et LSTM. Nous remarquons que l'algorithme XGBoost avec l'approche par classification performe mieux que les autres, suivi par l'algorithme LSTM avec l'approche par intervalles interdemandes.

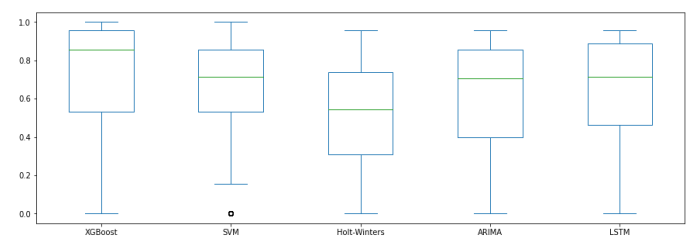


Figure 8 : F1-score de l'occurrence des ventes avec l'approche par classification, ainsi qu'avec l'approche par intervalles interdemandes en utilisant différents algorithmes

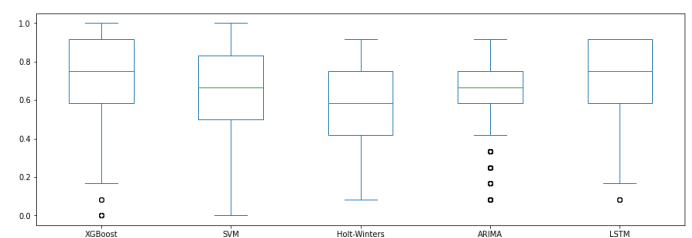
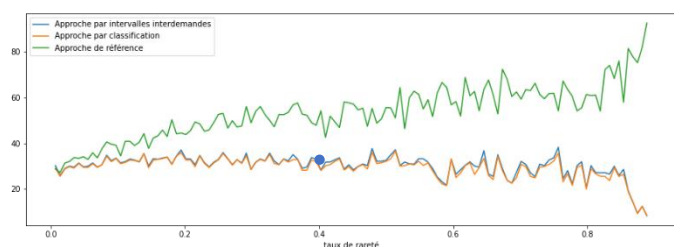


Figure 9 : Exactitude de l'occurrence des ventes avec l'approche par classification, ainsi qu'avec l'approche par intervalles interdemandes en utilisant différents algorithmes

### 5.1.2 Comparaison de la performance globale des deux approches (Étapes 1 et 2)

Nous proposons de comparer la performance de l'approche par classification et l'approche par intervalles interdemandes avec différents algorithmes à l'Étape 1, ainsi qu'à l'Étape 2. Pour chaque produit et avec chacune des deux approches, nous retenons la combinaison des algorithmes qui permet d'obtenir la sMAPE la plus faible.

Figure 10 illustre l'effet du taux de rareté (et donc de l'intermittence) sur l'erreur de prévision pour les deux approches (par classification ou par intervalles interdemandes) et de le comparer avec celle de l'approche de référence. Par exemple, le point bleu sur la figure représente la sMAPE moyenne sur tous les produits ayant un taux de rareté de 0,4 obtenue avec l'approche par intervalles interdemandes, sachant que pour chaque produit la meilleure combinaison d'algorithmes qui performe mieux respectivement à l'Étape 1 et à l'Étape 2 a été retenue. Aussi, pour chaque produit, la meilleure approche de référence est retenue, puis l'erreur moyenne sur tous les produits est affichée par la courbe verte.



**Figure 10 : Effet de l'intermittence sur l'erreur de prévision (sMAPE) obtenue avec l'approche par classification et par intervalles interdemandes**

Figure 10 montre clairement que l'approche par classification performe mieux que l'approche par intervalles interdemandes. Cependant, cela ne signifie pas que l'approche par classification est la meilleure pour tous les produits. Tableau 1 résume pour chaque approche et combinaison d'algorithmes (respectivement à l'Étape 1 et à l'Étape 2) le nombre de produits pour lesquels on obtient les meilleures prévisions avec l'approche et la combinaison d'algorithmes en question.

**Tableau 1. Nombre des produits (N) pour lesquels on obtient les meilleures prévisions avec une approche et combinaison d'algorithmes données**

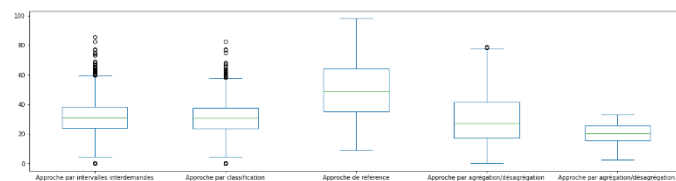
Approche	Combinaison d'algorithmes	N
Classification	XGBoost+ARIMA	634
	SVM+ARIMA	602
Intervalles Interdemandes	ARIMA+ARIMA	289
	ARIMA+Holt-Winters	201
	ARIMA+LightGBM	191
	Holt-Winters+Holt-Winters	95
	Holt-Winters+ARIMA	87
	LSTM+Holt-Winters	72
	LSTM+ARIMA	66

### 5.2 Approche par agrégation/désagrégation temporelle

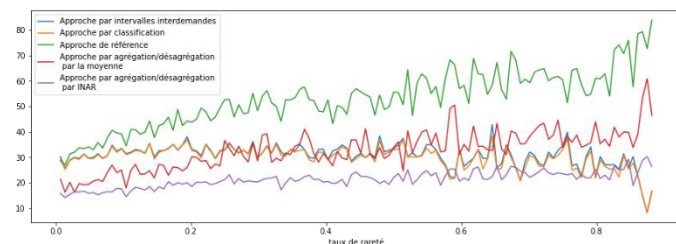
Dans cette section, nous procédons à une analyse des résultats de prévision à l'aide de l'approche d'agrégation et de désagrégation temporelle. Nous avons commencé par agréger les données sur une base trimestrielle dans le but de réduire l'intermittence, puis appliquer les modèles LSTM, ARIMA et Holt-Winters sur les séries temporelles agrégées. Pour obtenir des prévisions mensuelles, nous avons divisé par 3 pour avoir la moyenne, ou divisé par un coefficient prédit par le modèle

INAR. Finalement, nous retenons pour chaque produit le modèle de prévision (entre ARIMA, LSTM et Holt-Winters) qui performe le mieux.

La performance des approches par agrégation/désagrégation est comparée à celles des approches précédentes dans Figures 11 et 12. Premièrement, nous remarquons que les deux approches par agrégation/désagrégation performe mieux, en moyenne sur tous les produits, que les approches par classification et par intervalles interdemandes. Deuxièmement, d'après Figure 12, il est clair que l'approche utilisant un coefficient de désagrégation prédit par le modèle INAR est meilleure que tous les autres approches pour tous les produits, alors que celle utilisant la moyenne pour la désagrégation est supérieure aux autres approches seulement pour les produits à faible intermittence. On pourrait conclure donc que l'approche utilisant un coefficient de désagrégation prédit par le modèle INAR performe mieux que les autres approches, peu importe le niveau d'intermittence de la demande.



**Figure 11 : Erreurs de prévision (sMAPE) obtenues avec les trois approches hybrides et la méthode de référence**



**Figure 12 : Effet de l'intermittence sur l'erreur de prévision (sMAPE) obtenue avec les trois approches hybrides et la méthode de référence**

## 6 CONCLUSION

En étudiant le cas d'un distributeur situé au Québec et dont les clients sont des industriels, cet article propose une méthodologie de prévision des ventes dans un contexte B2B pour les produits à demande intermittente. Nous avons évalué l'efficacité de différentes méthodes de prévision avancées pour prévoir les ventes de produits à demande intermittente en utilisant l'historique des ventes réelles réalisées entre 2007 et 2019.

Les méthodes de référence n'ayant pas permis d'obtenir des prévisions assez précises, nous avons testé trois approches hybrides avec différentes combinaisons d'algorithmes. En retenant pour chaque produit la meilleure approche et la meilleure combinaison d'algorithmes, nous avons pu générer des prévisions de ventes plus précises et ainsi atténuer l'effet de l'intermittence de la demande. Ces résultats confirment l'intérêt des approches hybrides pour la prévision des ventes de produits à demande intermittente dans un contexte B2B. Une version étendue de cet article investiguera s'il existe une correspondance entre un type particulier de produits à demande intermittente et la méthode de prévision à privilégier.

## 7 REMERCIEMENTS

Les auteurs remercient MITACS et l'entreprise JSV Inc. pour le support financier.

## 8 REFERENCES

- Abolghasemi, Mahdi, Rob J Hyndman, Garth Tarr, and Christoph Bergmeir. 2019. 'Machine learning applications in time series hierarchical forecasting', *arXiv preprint arXiv:1912.00370*.
- Alwan, Layth C., and Christian H. Weiß. 2017. 'INAR implementation of newsvendor model for serially dependent demand counts', *International Journal of Production Research*, 55: 1085-99.
- Athanasopoulos, George, Roman A Ahmed, and Rob J Hyndman. 2009. 'Hierarchical forecasts for Australian domestic tourism', *International Journal of Forecasting*, 25: 146-66.
- Awad, Mariette, Rahul Khanna, Mariette Awad, and Rahul Khanna. 2015. 'Support vector machines for classification', *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*: 39-66.
- Babai, M. Zied, Mohammad M. Ali, and Konstantinos Nikolopoulos. 2012. 'Impact of temporal aggregation on stock control performance of intermittent demand estimators: Empirical analysis', *Omega*, 40: 713-21.
- Chen, Chao, Jamie Twycross, and Jonathan M Garibaldi. 2017. 'A new accuracy measure based on bounded relative error for time series forecasting', *PLOS ONE*, 12: e0174202.
- Chicco, Davide, Matthijs J Warrens, and Giuseppe Jurman. 2021. 'The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation', *PeerJ Computer Science*, 7: e623.
- Croston, John D. 1972. 'Forecasting and stock control for intermittent demands', *Journal of the Operational Research Society*, 23: 289-303.
- Feng, Yun-Wen, Jun-Yu Chen, Cheng Lu, and Shun-Peng Zhu. 2021. 'Civil aircraft spare parts prediction and configuration management techniques: review and prospect', *Advances in Mechanical Engineering*, 13: 16878140211026173.
- Ferhat, YUNA, and Burak ErKayman. 2021. 'A Genetic Algorithm-Based Model for Inventory Control in Intermittent Demands', *Avrupa Bilim ve Teknoloji Dergisi*: 696-701.
- Gourieroux, Christian, and Joann Jasiak. 2004. 'Heterogeneous INAR (1) model with application to car insurance', *Insurance: Mathematics and Economics*, 34: 177-92.
- Kourentzes, Nikolaos. 2013. 'Intermittent demand forecasts with neural networks', *International journal of production economics*, 143: 198-206.
- Kourentzes, Nikolaos, and George Athanasopoulos. 2021. 'Elucidate structure in intermittent demand series', *European Journal of Operational Research*, 288: 141-52.
- Lei, Ming, Shalang Li, and Qian Tan. 2016. 'Intermittent demand forecasting with fuzzy markov chain and multi aggregation prediction algorithm', *Journal of Intelligent & Fuzzy Systems*, 31: 2911-18.
- Liu, Jie, Lin Lin, Zhen Li, Hao Guo, and Yancheng Lv. 2020. 'Spare aeroengine demand prediction model based on deep Croston method', *Journal of Aerospace Information Systems*, 17: 125-33.
- Nikolopoulos, Konstantinos I., M. Zied Babai, and Konstantinos Bozos. 2016. 'Forecasting supply chain sporadic demand with nearest neighbor approaches', *International journal of production economics*, 177: 139-48.
- Nikolopoulos, Konstantinos, Aris A Syntetos, John E Boylan, Fotios Petropoulos, and Vassilis Assimakopoulos. 2011. 'An aggregate-disaggregate intermittent demand approach (ADIDA) to forecasting: an empirical proposition and analysis', *Journal of the Operational Research Society*, 62: 544-54.
- Novaković, Jasmina Dj, Alempije Veljović, Siniša S Ilić, Željko Papić, and Milica Tomović. 2017. 'Evaluation of classification models in machine learning', *Theory and Applications of Mathematics & Computer Science*, 7: 39.
- Petropoulos, Fotios, and Nikolaos Kourentzes. 2015. 'Forecast combinations for intermittent demand', *Journal of the Operational Research Society*, 66: 914-24.
- Skariah, Merin, and Chethamangalath Damodaran Suriyakala. 2022. 'Forecasting reservoir inflow combining Exponential smoothing, ARIMA, and LSTM models', *Arabian Journal of Geosciences*, 15: 1292.
- Spiliotis, Evangelos, Fotios Petropoulos, Nikolaos Kourentzes, and Vassilios Assimakopoulos. 2020. 'Cross-temporal aggregation: Improving the forecast accuracy of hierarchical electricity consumption', *Applied Energy*, 261: 114339.
- Spithourakis, Georgios P, Fotios Petropoulos, Konstantinos Nikolopoulos, and Vassilios Assimakopoulos. 2014. 'A systemic view of the ADIDA framework', *IMA Journal of Management Mathematics*, 25: 125-37.
- Syntetos, Aris A, John E Boylan, and JD Croston. 2005. 'On the categorization of demand patterns', *Journal of the Operational Research Society*, 56: 495-503.
- Syntetos, Aris A., Zied Babai, John E. Boylan, Stephan Kolassa, and Konstantinos Nikolopoulos. 2016. 'Supply chain forecasting: Theory, practice, their gap and the future', *European Journal of Operational Research*, 252: 1-26.
- Tan, Hui Xing, Nway Nway Aung, Jing Tian, Matthew Chin Heng Chua, and Youheng Ou Yang. 2019. 'Time series classification using a modified LSTM approach from accelerometer-based data: A comparative study for gait cycle detection', *Gait & posture*, 74: 128-34.
- Teunter, Ruud H, Aris A Syntetos, and M Zied Babai. 2011. 'Intermittent demand: Linking forecasting to inventory obsolescence', *European Journal of Operational Research*, 214: 606-15.
- Tripathy, Sarita, and Laxman Sahoo. 2020. "Improved Method for Noise Detection by DBSCAN and Angle Based Outlier Factor in High Dimensional Datasets." In *ICCCE 2019: Proceedings of the 2nd International Conference on Communications and Cyber Physical Engineering*, 213-21. Springer.
- Turkmen, Ali Caner, Yuyang Wang, and Tim Januschowski. 2019. 'Intermittent demand forecasting with deep renewal processes', *arXiv preprint arXiv:1911.10416*.
- Wang, Chen, Chengyuan Deng, and Suzhen Wang. 2020. 'Imbalance-XGBoost: leveraging weighted and focal losses for binary label-imbalanced classification with XGBoost', *Pattern Recognition Letters*, 136: 190-97.
- Willemain, Thomas R., Charles N. Smart, Joseph H. Shockor, and Philip A. DeSautels. 1994. 'Forecasting intermittent demand in manufacturing: a comparative evaluation of Croston's method', *International Journal of Forecasting*, 10: 529-38.
- Xu, Q., N. Wang, and H. Shi. 2012. "Review of Croston's method for intermittent demand forecasting." In *2012 9th International Conference on Fuzzy Systems and Knowledge Discovery*, 1456-60.
- Yacouby, Reda, and Dustin Axman. 2020. "Probabilistic extension of precision, recall, and f1 score for more thorough evaluation of classification models." In *Proceedings of the first workshop on evaluation and comparison of NLP systems*, 79-91.