

Prédiction du délai de vente des véhicules et optimisation du stock par apprentissage automatique

Hamid Ahaggach¹, Lylia Abrouk,² Eric Lebon ¹

² Laboratoire LIB , Université Bourgogne Europe, Dijon, France

prenom.nom@u-bourgogne.fr

¹ Syartec, Aix-en-Provence, France

hahaggach@Syartec.com, elebon@Syartec.com

Résumé

La gestion des stocks automobiles nécessite d'anticiper le délai de vente des véhicules afin de limiter les coûts d'immobilisation et la dépréciation. Cet article présente la démonstration d'un système d'aide à la décision fondé sur l'apprentissage automatique pour prédire le délai de vente d'un véhicule à partir de ses caractéristiques techniques et commerciales, puis recommander des décisions d'approvisionnement. Le problème est formulé comme une tâche de classification multi-classes, où le délai de vente est discrétisé en intervalles temporels. Le système proposé intègre un pipeline complet de prétraitement, un module de prédiction supervisée et un module décisionnel permettant d'estimer un stock cible compatible avec un objectif de rotation. La démonstration met en évidence l'intégration opérationnelle du modèle prédictif dans une interface interactive permettant d'explorer des scénarios de type what-if pour la gestion proactive des stocks.

Mots-clés

Classification supervisée, Sélection de caractéristiques, Optimisation décisionnelle; Système d'aide à la décision; Random Forest; Analyse prédictive; Gestion de stock automobile.

Abstract

Automotive inventory management requires anticipating vehicle time-to-sale in order to reduce holding costs and depreciation. This paper presents the demonstration of a machine-learning-based decision-support system designed to predict vehicle sales delay from technical and commercial attributes, and to support procurement decisions. The problem is formulated as a multi-class classification task, where time-to-sale is discretized into time intervals. The proposed system integrates a complete pipeline including data preprocessing, supervised prediction, and a decision module that estimates stock levels compatible with a target sales horizon. The demonstration highlights the operational integration of predictive models into an interactive application that enables users to explore what-if scenarios for proactive inventory management.

Keywords

Supervised classification; Feature selection; Decision optimization; Decision support system; Random Forest; Predictive analytics; Automotive inventory management.

1 Introduction

La rotation de stock constitue un enjeu central pour les concessionnaires et les plateformes de revente [4] : un véhicule immobilisé trop longtemps génère des coûts (financement, stockage, assurance), augmente le risque de décote et dégrade la trésorerie. À l'inverse, des stocks trop faibles entraînent des ruptures et des pertes d'opportunité commerciale. Dans ce contexte, anticiper le délai de vente d'un véhicule, dès son entrée en stock ou lors d'une décision d'achat, permet d'arbitrer entre disponibilité, marge et risque. Des travaux récents montrent que l'analyse prédictive et l'apprentissage automatique peuvent améliorer les décisions d'inventaire et réduire les coûts opérationnels dans l'industrie automobile [6, 8, 5].

Les délais de vente dépendent d'un ensemble de facteurs hétérogènes (marque, modèle, version, motorisation, couleur, saisonnalité, historique de stock), et les données opérationnelles sont souvent incomplètes, irrégulières et non stationnaires (évolution des préférences, politiques commerciales, contraintes d'approvisionnement). L'intégration de modèles d'intelligence artificielle et d'analytique avancée permet d'exploiter ces sources d'information hétérogènes afin d'améliorer la précision des prévisions et la gestion des stocks [3, 1].

Les approches de séries temporelles (ARIMA, Prophet, modèles séquentiels) supposent généralement des observations régulières et des historiques suffisamment longs par référence (produit, modèle, version). Or, dans les stocks automobiles, l'historique par référence fine (marque-modèle-version-options) est limité et irrégulier. Beaucoup de références apparaissent sporadiquement, rendant l'apprentissage de motifs temporels instable, et les ruptures d'approvisionnement introduisent des périodes sans observation. Des études récentes soulignent que dans ces contextes à demande intermittente, les méthodes d'apprentissage supervisé sur données tabulaires peuvent offrir une meilleure robustesse que les modèles temporels classiques [7, 2]. En

pratique, le problème se prête mieux à une formulation tabulaire supervisée : on prédit une étiquette de délai à partir de variables explicatives décrivant le véhicule et son contexte (date, stock, etc.). Cette approche permet : (i) d’exploiter des données multi-références dans un modèle global, (ii) d’intégrer naturellement des variables catégorielles, (iii) de produire des recommandations actionnables.

2 Problème et approche

On considère un concessionnaire qui achète et revend des véhicules. Chaque véhicule est décrit par un vecteur de caractéristiques x regroupant des informations techniques, commerciales et contextuelles (marque, modèle, version, motorisation, couleur, date d’estimation, niveau de stock, etc.). L’objectif est de prédire le délai de vente y , discrétisé en plusieurs classes :

$$y \in \{<3 \text{ m}, 3-6 \text{ m}, 6-9 \text{ m}, 9-12 \text{ m}, >12 \text{ m}\}.$$

Le premier objectif est donc de résoudre une tâche de classification supervisée :

$$\hat{y} = f_{\theta}(x),$$

où f_{θ} désigne un classifieur appris à partir de données historiques.

Le second objectif est décisionnel. Pour une référence donnée, le concessionnaire souhaite estimer un stock maximal recommandé n^* compatible avec un objectif de rotation, par exemple rester dans la classe <3 mois. Le principe consiste à faire varier la variable de stock dans le vecteur d’entrée afin d’identifier le seuil à partir duquel la classe prédite change.

La figure 1 présente l’architecture générale du système de démonstration. Celui-ci est structuré en trois composantes : un module de prétraitement des données, un module de prédiction du délai de vente, et un module d’optimisation décisionnelle. Les données manipulées sont de nature tabulaire et comprennent des variables catégorielles (marque, modèle, version, motorisation, couleur), des variables temporelles dérivées de la date d’estimation, ainsi qu’une variable d’action correspondant au nombre d’exemplaires en stock. Le prétraitement assure la validation des entrées, la construction de variables temporelles, l’encodage des variables catégorielles et la gestion des incohérences ou valeurs manquantes.

Le module de prédiction applique un classifieur supervisé au vecteur encodé. Plusieurs familles de modèles peuvent être envisagées, notamment les arbres de décision, les forêts aléatoires ou les méthodes d’ensemble. Dans notre prototype, le choix d’une Random Forest offre un compromis satisfaisant entre robustesse, simplicité d’usage et capacité à modéliser des interactions non linéaires entre variables. Le module décisionnel transforme ensuite la prédiction en recommandation d’approvisionnement. Pour une référence donnée, le système simule différentes valeurs de stock et observe l’évolution de la classe prédite. Une stratégie en deux temps est appliquée : une recherche grossière pour localiser une zone de bascule, puis une recherche fine pour

estimer précisément le seuil de réapprovisionnement. Le système peut ainsi fournir soit un nombre maximal de véhicules à acheter, soit une recommandation d’attente avant réapprovisionnement.

3 Démonstration du système

Afin d’illustrer l’intégration opérationnelle de l’approche proposée, nous avons développé un prototype interactif permettant d’explorer les prédictions et les recommandations de stock. L’objectif de la démonstration est de montrer comment un utilisateur métier peut renseigner les caractéristiques d’un véhicule, obtenir une estimation du délai de vente et analyser différents scénarios de réapprovisionnement.

L’interface du prototype a été conçue pour un usage décisionnel. Elle permet de saisir les principales caractéristiques d’un véhicule, d’afficher le délai de vente prédit et de consulter les recommandations associées. L’application propose également des vues complémentaires, notamment une vue de prédiction et une vue dédiée aux recommandations ou véhicules similaires. La restitution des résultats s’appuie sur une hiérarchie visuelle simple combinant zones de paramétrage, cartes de synthèse et tableaux interactifs. Cette organisation facilite l’exploration rapide des résultats par des utilisateurs non spécialistes de l’apprentissage automatique.

Lors de la démonstration, l’utilisateur sélectionne une configuration de véhicule à partir de ses caractéristiques principales et du stock courant. Le système retourne alors une classe de délai de vente prédite, accompagnée d’une recommandation de réapprovisionnement. L’utilisateur peut ensuite modifier la valeur du stock pour explorer des scénarios de type *what-if* et observer l’évolution de la recommandation. Ce scénario met en évidence l’intérêt du système pour l’aide à la décision dans un contexte de gestion proactive des stocks.

Le prototype est implémenté sous la forme d’une application web interactive (Figure 2). Afin de garantir une expérience fluide lors de la démonstration, le système met en cache les ressources du modèle ainsi que certains calculs intermédiaires. Cette stratégie limite les recalculs inutiles et permet de conserver les résultats lors des interactions successives avec l’interface.

4 Évaluation

L’évaluation du système repose sur un jeu de données opérationnel fourni par deux concessionnaires automobiles couvrant la période d’octobre 2013 à novembre 2023. L’ensemble de données comprend plus de 73 200 véhicules décrits par 34 attributs techniques et commerciaux, incluant notamment la marque, le modèle, la motorisation, l’année, ainsi que diverses informations liées au stock et à la commercialisation. Les données proviennent de deux entreprises distinctes et incluent à la fois des véhicules neufs (VN) et des véhicules d’occasion (VO) [4].

Les données ont été divisées aléatoirement en deux ensembles : 80% pour l’entraînement et 20% pour l’évalua-

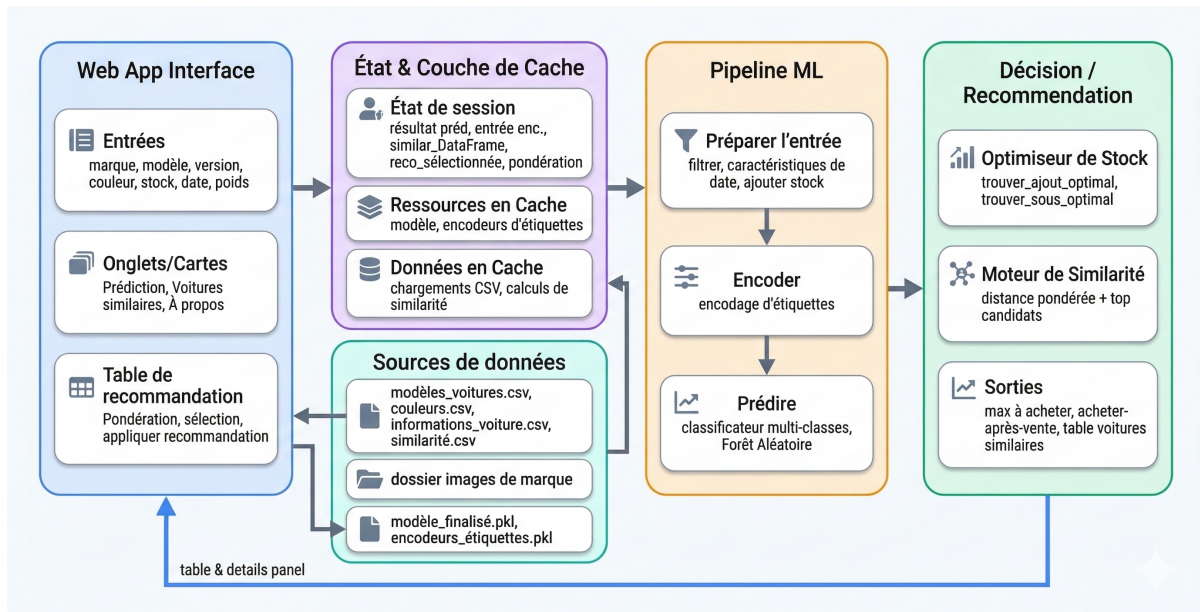


FIGURE 1 – Architecture générale du système de démonstration.

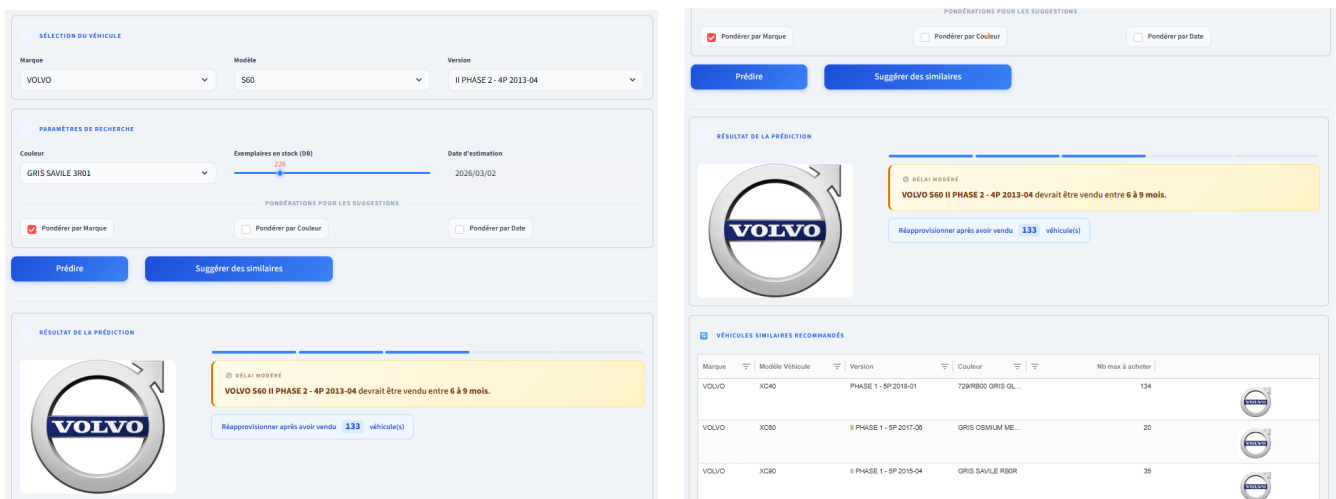


FIGURE 2 – Vues du prototype de démonstration : interface de prédiction et vue de recommandation.

tion. L'objectif du modèle est de prédire l'intervalle de temps pendant lequel un véhicule restera en stock avant d'être vendu. Les délais de vente sont discrétisés en plusieurs classes correspondant à des intervalles temporels (par exemple : 0-3 mois, 3-6 mois, 6-9 mois, 9-12 mois, >12 mois).

Plusieurs algorithmes d'apprentissage supervisé ont été évalués, notamment KNN, SVM, Decision Tree et Random Forest. Les hyperparamètres ont été optimisés à l'aide d'une recherche par grille (*GridSearch*) afin d'obtenir les meilleures performances possibles pour chaque modèle.

Les résultats expérimentaux montrent que le modèle Random Forest obtient les meilleures performances globales. Pour le premier concessionnaire, le modèle atteint une précision de 99.53% pour les véhicules neufs et 92.09% pour les véhicules d'occasion. Des résultats similaires sont observés pour le second concessionnaire, avec une précision

de 99.24% pour les véhicules neufs et 91.80% pour les véhicules d'occasion. Ces performances s'expliquent en grande partie par la capacité des forêts aléatoires à combiner plusieurs arbres de décision et à capturer des relations non linéaires complexes dans les données.

Les matrices de confusion montrent que la majorité des véhicules sont correctement classés dans leur intervalle de délai de vente. Les erreurs observées concernent principalement les classes adjacentes, notamment entre les intervalles 0-3 mois et 3-6 mois. Ce phénomène est cohérent avec la nature du problème : deux véhicules très similaires peuvent être vendus à quelques jours d'intervalle et être ainsi associés à des classes différentes.

Ces résultats confirment la pertinence de l'approche proposée pour la prédiction du délai de vente. Ils montrent également que les modèles d'ensemble, tels que Random Forest, constituent une solution robuste pour exploiter des données

tabulaires hétérogènes issues de contextes opérationnels.

5 Discussion

Les résultats expérimentaux mettent en évidence plusieurs éléments importants concernant la prédiction du délai de vente dans le secteur automobile. Tout d'abord, les performances obtenues montrent que les approches d'apprentissage supervisé sur données tabulaires sont particulièrement adaptées à ce type de problème. Contrairement aux modèles de séries temporelles classiques, qui nécessitent des historiques longs et réguliers par référence, les modèles tabulaires peuvent exploiter simultanément les caractéristiques de nombreux véhicules et généraliser plus facilement à de nouvelles configurations.

L'utilisation d'un modèle Random Forest s'avère particulièrement pertinente dans ce contexte. Les forêts aléatoires permettent de capturer des interactions complexes entre variables, tout en restant relativement robustes au bruit et aux données manquantes. Dans notre cas, elles offrent un bon compromis entre précision, stabilité et temps d'entraînement.

L'analyse des matrices de confusion révèle cependant certaines limites. Les erreurs de classification apparaissent principalement entre classes temporelles proches. Par exemple, un véhicule vendu en trois mois et un jour est classé différemment d'un véhicule vendu en deux mois et vingt-neuf jours, bien que ces situations soient très proches d'un point de vue métier. Ce phénomène est inhérent à la discrétisation du délai de vente en classes.

Une autre limite concerne la variabilité des marchés automobiles. Les préférences des consommateurs, les politiques commerciales ou les conditions économiques peuvent évoluer rapidement, ce qui peut entraîner une dérive des distributions de données au cours du temps. Dans un contexte opérationnel, il est donc nécessaire de réentraîner régulièrement les modèles afin de maintenir leur performance.

Malgré ces limites, les résultats montrent que l'intégration d'un modèle prédictif dans un système interactif peut constituer un outil d'aide à la décision pertinent pour les concessionnaires. En combinant prédiction et simulation de scénarios, il devient possible d'explorer différents niveaux de stock et d'évaluer leur impact potentiel sur la rotation des véhicules.

6 Conclusion

Dans cet article, nous avons présenté la démonstration d'un système d'aide à la décision dédié à la prédiction du délai de vente et à l'optimisation des stocks automobiles. L'approche proposée repose sur un modèle d'apprentissage supervisé exploitant les caractéristiques techniques et commerciales des véhicules afin d'estimer leur délai de vente sous forme d'intervalles temporels.

Les expérimentations réalisées sur un jeu de données réel comprenant plus de 73 000 véhicules ont montré que les modèles d'ensemble, en particulier Random Forest, permettent d'obtenir des performances élevées pour cette tâche. Les résultats confirment que l'utilisation de données

tabulaires hétérogènes constitue une approche efficace pour modéliser les dynamiques de vente dans le secteur automobile.

Au-delà de la performance prédictive, l'un des apports principaux de ce travail réside dans l'intégration du modèle dans un prototype interactif permettant d'explorer des scénarios décisionnels. L'utilisateur peut ainsi analyser l'impact de différents niveaux de stock sur le délai de vente prédit et obtenir des recommandations opérationnelles pour la gestion des inventaires.

Ces résultats ouvrent plusieurs perspectives de recherche. À court terme, l'intégration de variables économiques supplémentaires telles que le prix, les promotions ou la marge commerciale pourrait améliorer la qualité des prédictions. À plus long terme, des approches hybrides combinant apprentissage automatique et connaissances métier, par exemple sous forme d'ontologies ou de règles décisionnelles, pourraient permettre d'améliorer l'interprétabilité des recommandations et leur adoption dans des contextes industriels.

Références

- [1] Advancements in sales forecasting : A critical evaluation of machine learning approaches. *International Journal of Engineering Trends and Technology*, 2024.
- [2] Demand forecasting in the automotive industry : A systematic literature review. *MDPI Systems*, 2024.
- [3] Machine learning-based demand and pricing optimization for a resilient automotive supply chain. *IJRASET*, 2024.
- [4] Hamid Ahaggach, Lylia Abrouk, Sebti Fofou, and Eric Lebon. Predicting car sale time with data analytics and machine learning. In *IFIP International Conference on Product Lifecycle Management*, pages 399–409. Springer, 2022.
- [5] So Fukuhara, Abdallah Alabdallah, Nuwan Gunasekara, and Slawomir Nowaczyk. Bridging forecast accuracy and inventory kpis : A simulation-based software framework. *arXiv preprint arXiv :2601.21844*, 2026.
- [6] Pranav et al. Gaikwad. Predictive logistics management of car sales based on machine learning algorithm for supply chain. 2024.
- [7] Cornelius Hake, Jonas Weigele, Frederik Reichert, and Christian Friedrich. Evaluation of artificial intelligence methods for lead time prediction in automotive production. *arXiv preprint arXiv :2501.07317*, 2025.
- [8] Manoj Omprakash. Optimizing demand forecasting and inventory management with ai in the automotive industry. Master's thesis, LUT University, 2024.