# Project: Implementing Machine Learning tools and/or techniques in Churn prediction using Cartrack Data.

1. What is the focus and aim of your review? Who is your audience?

**Focus:** ML approaches for predicting churn across business to business (B2B) and individual clients or business to client (B2C), grounded in Cartrack's context.

**Aim:** Synthesize 2019–2025 evidence and translate into a deployable pipeline (features, methods, metrics, governance) to reduce churn and increase profitable retention.

**Audience:** Cartrack/Karooooo managers, analytics leadership, and MSc assessors. Cartrack's current scale (2.4 M subscribers) underscores impact. (Karooooo, N.D.)

2. Why is there a need for your review? Why is it significant?

At Cartrack's scale, even small gains materially affect Annual Recurring Revenue (ARR). A proactive, ML-driven program shifts retention from reactive "save after notice" to early risk + timing + treatment targeting.

3. What is the context of the topic or issue? What perspective do you take? What framework do you use to synthesise the literature?

**Context:** Telematics churn focuses on full account exits. Signals live in subscriptions, usage by client, device health and repair stats.

**Perspective:** Practical use: models that are ready to deploy, avoid data leakage, keep costs in mind, and balance accuracy, reliability (calibration), explainability, and financial return.

**Synthesis framework:** PRISMA (a structured search and screening method) for source screening. (Akl et al, 2024)

4. How did you locate and select sources for inclusion in the review?

**Databases:** Google Scholar is always my go too. (Google, N.D.)

**Inclusion:** Only papers released after 2019

5. How is your review structured?

Introduction & scope

Method (PRISMA - explain)

Algorithms & models' basics

Features & data design

Evaluation & validation

Explainability & governance

Gaps & agenda for telematics

Implications for Cartrack

Conclusion

6. What are the main findings in the literature on this topic?
- Machine learning (ML) and Deep learning (DL) are strong when predicting churn. Specifically on tabular data. (Hasan 2025) (Imani, Beikmohamadi, & Arabnia, 2025)
- Calibration + interpretability matter for deployment (Văduva, Oprea, Niculae, Bâra, & Andreescu, 2024)
- More variables like tickets, client behavior and transaction history help identify early signs of churn. (Bolaji, Chukwuma-Eke, Balogun, & Ogunsola, 2023)

7. What are the main strengths and limitations of this literature?
- Some papers are not peer reviewed like (Imani, Beikmohamadi, & Arabnia, 2025)
- Though we can see many tips to improve the churn models like shown above where more variables improve models and ML and DL are better churn predictors. This literature cannot help in the domain expertise needed on Cartracks data and abnormalities. Like Churn being pinged when a client just owner changed their contracts to a different company they own.

8. Are there any discrepancies in this literature?
- Not all papers agree on which model is best to use.
- The papers also use different evaluation techniques from ROC-AUC to calibration/profit.

9. What conclusions do your draw from the review? What do you argue needs to be done as an outcome of the review?
- I will use gradient-boosted trees with SHAP and probability calibration as my base model. With multiple models to compare against being built after this one.
- I need to make sure my training data is clean and accurate. Run it through finance to make sure I am accurately flagging all churn.
- B2B and B2C might require different models due to having different variables for prediction.

References:

Karooooo (N.D.) Welcome to Karooooo Available from: https://karooooo.com/ [Accessed 15 August 2025]

Akl, E. et al. (2024) Extension of the PRISMA 2020 statement for living systematic reviews (PRISMA-LSR): checklist and explanation. Available from: https://www.bmj.com/content/387/bmj-2024-079183 [Accessed 14 August 2025]

Google (N.D.) Google Scholar. Available from: https://scholar.google.com/ [Accessed 14 August 2025]

Hasan Mamun, H. (2025) "ADVANCEMENTS IN MACHINE LEARNING FOR CUSTOMER RETENTION: A SYSTEMATIC LITERATURE REVIEW OF PREDICTIVE MODELS AND CHURN ANALYSIS", Journal of Sustainable Development and Policy, 1(01), pp. 250–284. Available from: https://jsdp-journal.org/index.php/jsdp/article/view/11 [Accessed 15 August 2025]

Imani, M., Beikmohamadi, A., & Arabnia, H. (2025) Customer Churn Prediction: A Review of Recent Advances, Trends, and Challenges in Conventional Machine Learning and Deep Learning. Available from: https://www.preprints.org/manuscript/202503.1969/v2 [Accessed 15 August 2025]

Văduva, A., Oprea, S., Niculae, A., Bâra, A., & Andreescu, A. (2024). Improving Churn Detection in the Banking Sector: A Machine Learning Approach with Probability Calibration Techniques. Available from: https://www.mdpi.com/2079-9292/13/22/4527 [Accessed 15 August 2025]

Bolaji, A., Chukwuma-Eke, E., Balogun, E., & Ogunsola, K. (2023). Improving Customer Retention Through Machine Learning : A Predictive Approach to Churn Prevention and Engagement Strategies. Available from: https://www.researchgate.net/publication/390113580_Improving_Customer_Retention_Through_Machine_Learning_A_Predictive_Approach_to_Churn_Prevention_and_Engagement_Strategies [Accessed 15 August 2025]