Machine Learning for Churn Prediction in Telematics for Cartrack: A Literature Review

Contents

Introduction, Scope & Research Questions

Churn prediction offers potential retention strategies using reactive protocols before losing clients. The goal is to demonstrate how machine learning (ML) can help with that stepping stone. At Cartrack there is a current scale of over 2.4 million subscribers (Karooooo, N.D.), even small improvements in churn prediction accuracy can influence Annual Recurring Revenue (ARR).

This review focuses on the application of ML techniques for predicting churn in both business-to-business (B2B) and business-to-consumer (B2C) contexts, grounded in Cartrack's operational environment. The separation of the two contexts is due to the way we handle businesses and consumers. We will go into that further in a later section. The aim is to pull together clean Cartrack internal data from 2019–2025 to identify effective features, algorithms, metrics, and governance approaches for building a deployable churn prediction pipeline at Cartrack. The audience for this review includes Cartrack's analytics leadership, senior managers in retention and product, and university of Essex academic assessors. The relevance to the organisation is clear: with a high-volume, subscription-based model, any delay in identifying high-risk accounts has measurable financial consequences.

Research questions (RQ) to assist the literature review:

- RQ1: What machine learning algorithms perform best for churn prediction in telematics?
- RQ2: Which feature engineering and data design strategies are most effective in churn prediction?
- RQ3: How do studies address evaluation, calibration and explainability in churn models?

Method (PRISMA)

This review follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology. This is important as it allows the methods of finding supporting literature to be repeated by anyone (Akl et al, 2024). The literature found was sourced from Google Scholar (Google, N.D.), ensuring coverage of peer-reviewed journals and reputable preprints from 2020 onwards. Combining the two increases criticality and avoiding outdated work. Important factors when building ML models for production in a fast-paced business environment.

Inclusion criteria for literature:
- Published between 2020–2025.
- Focus on ML/ Deep learning (DL) approaches on churn prediction.

Exclusion criteria:
- Work published before 2020, unless directly referenced in included papers.
- Studies not accessible to the public without paywalls.

Algorithms & Models' Basics (RQ1)

Across the literature that was reviewed throughout this project, three main model types emerged: logistic regression, gradient-boosted tree, and DL model types.

- Logistic regression remains common in basic churn studies, due to it's easy to learn background mathematics and ease of deployment. Limitations for logistic regression is its performance is often limited when modelling complex, non-linear relationships (Mamun, 2025).

- Gradient-boosted tree algorithms such as XGBoost and LightGBM consistently outperform linear methods on tabular datasets, the main format the data in this project will be, due to their ability to model feature interactions and handle heterogeneous variable types (Mamun, 2025; Imani et al., 2025).

- Deep learning approaches show competitive performance in domains with sequential or unstructured behavioural data. However, in telematics churn prediction, DL's gains over tree ensembles are inconsistent unless enriched sequential data (e.g., event logs, time-series device metrics) is available (Imani et al., 2025).

Văduva et al. (2024) emphasise the importance of probability calibration, especially in operational contexts where business decisions depend on accurate probability thresholds. Without calibration, even high-AUC models may misclassify risk tiers, leading to inefficient retention interventions. In the context of this project, probability calibration is used instead of explaining the churn percentage for a client as "70% chance of churn" for a specific client. The output needs to rather read "Around 7 out of 10 similar customers churned with homogeneous variables".

Summary table to compare models:

Model Type	Strengths	Limitations	Best Use Case
Logistic Regression	Easy to understand and explain, fast to train and deploy.	Struggles with complex, non-linear relationships.	Good baseline model when transparency is more important than raw accuracy.
Gradient-Boosted Trees	High accuracy on tabular data, works well with mixed and missing values, models interactions well.	Needs tuning, harder to explain without tools like SHAP (SHapley Additive exPlanations). Method of explaining how each feature in the model contributes to a specific prediction)	Strong choice for churn prediction on Cartrack's structured telematics data.
Deep Learning	Handles complex, sequential, or unstructured data; can capture hidden patterns.	Needs large datasets; risk of overfitting; slower to train.	Best when there's rich behavioural or time-series data, like detailed event logs.

Features & Data Design (RQ2)

Features and data design are selected to provide the models with the parameters (settings) and data (client history) they will use to function. The literature highlights feature selection as a crucial choice in churn prediction (Bolaji et al., 2023). For example, at Cartrack, features that may affect churn include:

- **Subscription and contracts**: Invoice amounts and contract start dates.
- **Device health**: Repairs requested by the client and repairs complete by Cartrack.
- **Usage behaviour**: Login frequency into Cartrack applications like fleet.

Bolaji et al. (2023) found that adding feature selections that show behaviour and transaction history of clients significantly improved model performance for early churners. Allowing early detection to attempt prevention of the official churn or bulk cancelation from beginning. The issue comes with incorrect or missing data. Which can occur for many reasons such as admin error, client error or database errors. With issues in data increasingly inevitable as the company size grows, adding only "perfect" data is also not an option. As more data can lead to improved model results compared to less data given to the model (Imani et al. 2025). Selecting the correct features will take model performance comparisons and knowledge on the department. Combining key features with using snapshot tables to allow consistent retraining and evaluation throughout (IBM, 2021).

Though no data is yet being handled. Ethics to be aware of is we will be using client's personal data. (Republic of South Africa, 2013) Hence the model will avoid using and showing personally identifiable information like IDs, company names etc. Instead use internal user ID and contract IDs throughout. Since all data is internal and willingly given by the client. The main objective is not presenting any data publicly or to any Cartrack employee not involved in the project.

Evaluation & Validation (RQ3)

Evaluation metrics vary widely in the literature, reflecting different datasets or goals of the model/company change the sort of evaluation method used. ROC-AUC is the most common, but in retention contexts, it can overstate performance by weighting all thresholds equally (Mamun, 2025). Precision@K, the precision within the top K predicted churners, is favoured where resources for interventions are limited, as the model will only show the top k most likely to churn. PR-AUC offers a more realistic view of classifier performance on imbalanced datasets, which is typical in churn (Văduva et al., 2024).

Văduva et al. (2024) argue strongly for integrating probability calibration into the evaluation process, particularly when downstream actions have financial implications. Cost curves and profit-based metrics are also used in some studies to explicitly tie model outputs to revenue impact (Mamun, 2025). These can help show the business how the model can potentially save the company money by using the model.

Validation measurements need to take place. With churn data, client events unfold over time. Randomly splitting the data into train/test sets can cause a mix up of time lines. To avoid that, Forward-chaining (time-based) cross-validation can mirror real-world deployment. (Imani et al., 2025).

Explainability & Governance (RQ3)

Explainability is key for both stakeholders and regulatory compliance. The trust of stakeholders is integral as they will then push managers to make sure the model is used in production. Compliance needs to be met for both auditors and government requirements. Trust with stakeholders can be achieved using SHAP which as I explained above is a method of explaining how each feature in the model contributes to a specific prediction. Showing stakeholders the variables that help the model make the predictions they do and thus explain the final cost curves and profit-based metrics.

Governance extends beyond explainability to include reproducibility, version control, and documentation. This can all be done in GitHub, a typical Cartrack audit approved method, using the CRISP-DM process. The CRISP-DM process supports these elements by defining phases for business understanding, data preparation, modelling, evaluation, and deployment (IBM, 2021). Văduva et al. (2024) stress the value of aligning governance with model monitoring, ensuring that calibration and performance do not degrade unnoticed.

Gaps & Agenda for Telematics

While ML and DL approaches for churn prediction are well-researched, I have found several gaps in the literature for telematics specific domain:

1. **Calibration in the real world**: While calibration is discussed (Văduva et al., 2024), such as how calibration improves model prediction trustworthiness, the process or cut offs are under documented. Addressing this gap will take trial and error. If the model predicts a client has 70% chance of leaving, The models requires calibrations to help the audience to understand exactly what that means. SHAP and domain knowledge should help explain what 70% chance means exactly. This will ensure we do not mislead anyone with the predictions.

2. **B2B vs B2C segmentation**: Most studies treat business and individual clients homogeneously. Yet, from domain experience they are not treated the same. Billing and behaviour differ. This is why I will keep them separate to bridge this gap and rather create two different models. This will also lessen bias as larger companies may outshine top clients due to size of churn. Potentially starting with the context that affects the company more, B2B churn.

Implications for Cartrack

The reviewed literature provides a clear, evidence-based roadmap for how Cartrack can implement churn prediction models. The goals for the models are to be operationally efficient and financially beneficial. Across studies, gradient-boosted trees consistently emerge as the most suitable base model for tabular telematics data (Mamun, 2025; Imani, Beikmohamadi, & Arabnia, 2025), aligning directly with Cartrack's structured big datasets.

Bolaji et al. (2023) highlight that multiple columns/factors from multiple business departments improve early churn detection. For Cartrack, this means more data may lead to better predictions. Insight from multiple departments like finance, sales and debtors can all help in understanding factors that go into the loss of clients, as each client is unique. Some clients may not have been happy with our invoice methods as told by finance and shown by invoice types changing month to month. Some clients may have struggled onboarding new vehicles onto our system as described by sales. Some clients may have been in arrears as explained by the debtor's department. All cases that could have caused the churn of a client. With many more factors like systems department noticing an influx of tickets or repairs department noticing multiple repairs from a single client. All signs that could be used to predict churn for a future customer. Also giving an opportunity to see where our shortfalls are as a company.

Văduva et al. (2024) stress the importance of probability calibration when deploying models in decision-making workflows. For Cartrack, calibrations ensure the model output is understood and the right calls are made for company growth. No wasted time on smaller clients with small chance of loss but rather the bigger giants who provide the highest revenue for the company and who show clearer signs of churning.

This involves using SHAP to explain the exact factors that went into the churn prediction. Enabling the business to understand the exact drive into high churn. This could come down to a simple factor like companies going into arrears. Though we could use that to implement a solution to assist companies short on cash rather than loose them fully. Or the SHAP analysis could explain factors we never took into account before. Like our invoice system not being up to standard.

In summary, the literature fits Cartrack's needs well. They set a fantastic base for Cartrack to begin research on a potential positive financial impact and in many cases a more important factor of customer satisfaction.

Conclusion

The review critically examined recently released literature on ml approaches to churn prediction in telematics, framed by three research questions.

RQ1: Gradient-boosted tree consistently outperforms linear models on structed churn data and are well suited to Cartrack's environment.

RQ2: Behavioural, transactional, and device-health feature are the most effective for improving prediction accuracy. More columns are known to strengthen performance. Missing or noisy data should be investigated in an exploratory data analysis before hand to avoid bias or unreliability.

RQ3: ROC-AUC is positively widely reporting. SHAP and CRISP-DM are best for clear paths to model explainability to the audience and sustaining performance over time.

In conclusion, the current state of research provides a strong technical foundation for Cartrack to develop churn prediction models. While gaps and further investigation are highlighted to promote accurate and strategic companywide benefits.

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]

Mamun, M.N.H. (2025) ADVANCEMENTS IN MACHINE LEARNING FOR CUSTOMER RETENTION: A SYSTEMATIC LITERATURE REVIEW OF PREDICTIVE MODELS AND CHURN ANALYSIS. Available from: https://jsdp-journal.org/index.php/jsdp/article/view/11 [Accessed 15 August 2025]

Imani, M., Joudaki, M., Beikmohamadi, A., & Arabnia, H.R. (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.-G., Oprea, S.-V., Niculae, A.-M., Bâra, A., & Andreescu, A.-I. (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]

Adekunle, B.I., Chukwuma-Eke, E.C., Balogun, E.D., & Ogunsola, K.O. (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]

IBM (2021) CRISP-DM Help Overview. Available from: https://www.ibm.com/docs/en/spss-modeler/saas?topic=dm-crisp-help-overview [Accessed 15 August 2025]

Republic of South Africa (2013) super crisp for reconciliation Available from: https://www.gov.za/sites/default/files/gcis_document/201409/3706726-11act4of2013protectionofpersonalinforcorrect.pdf [Accessed 1 September 2025]

Feedback

72% (Distinction)

Thank you for your submission of your literature review, which ideally builds on the draft and formative feedback provided (if you have submitted one for comment) and on the materials covered in the literature review unit.

As you know, the assignment is marked according to a few components, such as your overall understanding of the topic and a clear outline of your own research, the degree of coverage of the current debate, and the quality and structure of the LR itself.

I think you did an overall satisfactory job with your LR, with some areas of possible improvement. Below you can find some specific comments.

Knowledge and Understanding

The word demonstrates ability to provide a critical discussion. You express your learning well. Your ideas are well-developed and coherent, although not always completely clear.

Critical discussion & Contribution

The review addressed especially well gaps and implications. Offering some additional insight in the conclusive section would have been even more convincing.

Moreover, additional academic literature could have been consulted to enhance the critical depth.

Structure and Presentation

The literature consulted demonstrates evidence of reading in the discipline. You structured your review well, with a good introduction, an attempt to analyse the literature, an argument, and a conclusion.

Harvard System has been overall correctly used

Please keep those points in mind for your final project.