**Name: Dheekonda. Murali Venkata Sai Krishna**

**Student Id: A00047378**

**Student mail-id: dheekonm@roehampton.ac.uk**

# Predicting Customer Churn in the Telecom Industry Using Machine Learning: A Literature Review

# Table of Contents

# Abstract

Customer churn forecasting is crucial to telecom operators that want to lower customer loss and raise retention. Machine learning allows for identifying the at-risk customers accurately based on the data about their demographics, the usage of services, contracts and billing. The ensemble approaches were especially capable of detecting complex patterns, whereas such methods of class imbalance lessen the identification of churning instances like synthetic oversampling. Having the predictive models together with the segmentation permits greater focus on the retention strategies, and interpretability techniques provide insights that are straightforward and aspirational. This literature review provides an overview of major knowledge developments in these domains by providing best practices in the development of accurate, fair and business-ready churn prediction systems. The review of the findings will help in formulating effective solutions that will result in better decisions and customer loyalty.

# 1. Introduction

Churn of customers is the telecom term utilized to explain the loss of clients; this impacts the revenues and long-term profitability directly. The prediction of churn is important to a business because it is usually expensive to obtain new customers as compared to retaining them. Machine learning has greatly helped in identifying at-risk customers with rich data about the customers including demographics, service usage, contracts and billing history. In binary classification, class imbalance and interpretability of results prove to be complex issues in the task, which is a binary classification problem at hand. This review is a summary of recent developments in ensemble modelling, imbalance management, segmentation and explainability which may guide the best practice in future work to be applied in reality.

# 2. Literature Review

According to Alotaibi and Haq (2024) testify that XGBoost, LightGBM, and Random Forest are the most efficient ensemble learning approaches to implementing customer churn prediction in the telecommunications sector because they work better than single classifiers in terms of accuracy and recall. They used a large portion of preprocessing, feature encoding and hyperparameter tuning using the IBM Telco Customer Churn dataset to reach an accuracy of about 80% and a recall of 0.72. Their results demonstrate the necessity to unite the model optimisation with robust data preparation in order to enhance retention strategies.

According to Wu et al. (2021) propose an integrated churn prediction and customer segmentation framework and provide that the integration of segmentation with churn risk leads to the design of more specific retention campaigns. The method they use is able to not only estimate the probability of churn but it also classifies customers into exploitable segments by behavioural and demographic groups. The two-skinned tactics enable automation of offering, pricing and improvement of services to be customised to suit each segment, optimising marketing resources. The framework makes interventions more precise and increases the return on investment resulting in retention activities by associating churn probabilities based on segment-specific characteristics. The operational benefit of integrating predictive analytics with customer relationship management systems and the ability to receive direct benefit from delivered customer relations in business processes, as shown in the study is another consideration.

According to Suguna et al. (2025) have worked on a longstanding problem of class imbalance in churn datasets, which can be very skewed i.e., there will be significantly fewer churners than the retained ones. Such an imbalance may lead to bias towards the majority class that will make models focus to reflect on them, which will lead to the poor detection of real churners. In response to this, they used SMOTE (Synthetic Minority Oversampling Technique), hybrid resampling methods, such as SMOTE (combined with undersampling) and ensemble classifiers, such as XGBoost and Random Forest. Not only did these methods expand the proportion of the churn cases in the training data, but they also retained key decision boundaries, greatly enhancing the sensitivity and recall of the models to the minority-class churners. In their experiments, they showed that these preprocessing methods can be added to modelling without a precision lost, resulting in a more even-handed performance per metric. It is important to note that the efficient management of imbalances is required to prevent bias on the majority classes and promote the possibility of maintaining high-value retention opportunities without overcoming them.

According to Noviandy et al. (2024) paid attention to the interpretability of the model, where a model-agnostic SHAP (Shapley Additive exPlanations) algorithm was applied in order to explain the global and local importance of features in churn prediction models. Their method gave a general sense of the most important drivers of churn in general, like contract type, tenure and per-month charges as well as being able to provide personalised detail about individual customers and how much each feature contributed to that prediction. Such a bi-fold view will make each business team able to identify interventions that are effective in general, as well as interventions that are specific to each subscriber's risk profile. The study highlighted that feature impact transparency helps to engender trust in decision-makers, establishes cooperation between technical and business teams and makes the adoption of automated predictions to customer relationship management systems more likely.

Taken together, these papers suggest that the most promising churn prediction systems in telecoms unite good modelling, of course, sound preprocessing, class imbalancing control, and actionability and explainability, achieving both good predictive qualities and meaningful actionability.

# 3. References


Muteb Zarraq Alotaibi and Mohd Anul Haq (2024). Customer Churn Prediction for Telecommunication Companies using Machine Learning and Ensemble Methods. *Engineering Technology & Applied Science Research*, 14(3), pp.14572–14578. doi: https://doi.org/10.48084/etasr.7480.

Suguna, R., Prakash, J.S., Pai, H.A., Mahesh, T.R., Kumar, V.V. and Yimer, T.E. (2025). Mitigating class imbalance in churn prediction with ensemble methods and SMOTE. *Scientific Reports*, [online] 15(1). doi: https://doi.org/10.1038/s41598-025-01031-0.

Teuku Rizky Noviandy, Ghalieb Mutig Idroes, Hardi, I., Mohd Afjal and Ray, S. (2024). A Model-Agnostic Interpretability Approach to Predicting Customer Churn in the Telecommunications Industry. *Infolitika Journal of Data Science*, 2(1), pp.34–44. doi: https://doi.org/10.60084/ijds.v2i1.199.

Wu, S., Yau, W.-C., Ong, T.-S. and Chong, S.-C. (2021). Integrated Churn Prediction and Customer Segmentation Framework for Telco Business. *IEEE Access*, 9, pp.62118–62136. doi: https://doi.org/10.1109/access.2021.3073776.