The Scientific Temper (2024) Vol. 15 (4): 3272-3280
E-ISSN: 2231-6396, ISSN: 0976-8653
Doi: 10.58414/SCIENTIFICTEMPER.2024.15.4.36
https://scientifictemper.com/

RESEARCH ARTICLE

Multi-model telecom churn prediction

S. Bhuvaneswari[1*], A. Nisha Jebaseeli[2]

Abstract

Customer turnover is likely to be significant in the telecom business due to its dynamic and competitive nature. Traditional measures of performance are inadequate in such a fluid environment to accurately portray organisational objectives. The reason behind this is because the performance measurements are not in line with the company goals. A multi-model telecom churn prediction (MMTCP) with minority upliftment techniques is presented in this work. It can handle data imbalance successfully and has a loss function that separates loss into two parts: loss due to incorrect prediction and loss due to unavoidable loss. First, utilising a set of training data and a number of diverse base learners, MMTCP generates predictions at the first level; second, using these predictions as a starting point, it uplifts the minority in the model. Gradient-boosted trees and naïve bayes make up the first stage, while one-class SVM is the basis of the second combiner stage. As compared to both current classifier models and the state-of-the-art churn prediction methods from literature, the experimental findings suggest that the MMTCP model exhibits 1 to 7% greater churn prediction levels and 1.3 to 1.7 times decreased loss levels.

Keywords: Churn prediction, Ensemble modeling, Gradient boosted trees, Multi-model stacking, Upliftment.

INTRODUCTION

The significant growth in the use of mobile devices is undeniable. The Socio-Economic Council of the United Nations anticipates that there will be 7.7 billion internet users in 2019, with about the same number of mobile phones (Ahmed and Maheswari, 2017). Penetration levels ranging from 97% to 100% have been seen in several emerging nations. The fact that there are more mobile phone connections than individuals shows how the market remains stagnant. This makes breaking into new markets extremely challenging, as customer churn occurs since the only way to recruit new consumers is to force them to switch operators. It is five to six times more expensive to acquire new consumers than it is to retain current (Amin *et al*., 2018). Keeping consumers happy and coming back for more is another great way to grow your client base. The focus of subscription-based businesses is on holding on to current customers rather than trying to attract new, less likely-to-churn clients. When a consumer abandons a company in favor of a rival, this phenomenon is known as customer churn. Customer churn is bad because it causes you to lose customers, which in turn reduces your opportunity costs and makes others talk bad about your business. While the former might be quantified, the latter is intractable. Consequently, attrition prediction is crucial for any business..

Campaigns are typically launched after churner prediction. To decrease the likelihood of churn, these initiatives provide incentives to customers who may otherwise churn(Amin and Anwar *et al*., 2016). These deals typically come with free upgrades or discounts and are contingent on meeting certain requirements. The corporation gets charged for the expense of the initiative regardless of its kind. Therefore, it is essential to offer reliable forecasts. If the forecasts are wrong, the business might lose money. Churn prediction methods have a big problem with the churn data's inherent data imbalance. If one class has a very high number of instances in comparison to the other classes, the data is considered unbalanced. The

[1]Department of Computer Applications, Annai College of Arts and Science, Affiliated to Bharathidasan University, Kovilacheri, Kumbakonam, Tamil Nadu, India.

[2]Department of Computer Science, Centre for Distance and Online Education, Bharathidasan University, Tiruchirapalli, Tamil Nadu. India.

***Corresponding Author:** S. Bhuvaneswari, Department of Computer Applications, Annai College of Arts and Science, Affiliated to Bharathidasan University, Kovilacheri, Kumbakonam, Tamil Nadu, India, E-Mail: drbhuvaneswari@outlook.com

How to cite this article: Bhuvaneswari, S., Jebaseeli, A.N. (2024). Multi-model telecom churn prediction. The Scientific Temper, **15**(4):3272-3280.

Doi: 10.58414/SCIENTIFICTEMPER.2024.15.4.36

Source of support: Nil

Conflict of interest: None.

Received: 05/10/2024 **Accepted:** 19/11/2024 **Published** : 20/12/2024

former represents the dominant group, while the latter represents the underrepresented ones. After looking at the numbers from the telecom sector, it's clear that there's a major imbalance: there are a lot more non-churners than churners. Therefore, the data associated with churners is a small minority, but the data associated with non-churners constitutes a large majority. Biassed training of classifiers is a significant consequence of imbalance, which results in effective prediction of the majority class and low prediction levels for the minority class. To demonstrate high prediction levels for both the majority and minority classes, a churn prediction model must be able to handle minority classes successfully(Amin and Shah *et al*., 2018). A churn prediction model called Multi-Model Telecom Churn Prediction (MMTCP) is introduced in this study. We propose a loss function that breaks down loss into its two main components—unavoidable loss and loss owing to incorrect predictions—rather than treating prediction losses as a single cost or loss. Through comparisons with both current classifier models and state-of-the-art churn prediction models published in the literature, it was determined that the suggested MMTCP model achieves superior performance in terms of both prediction rates and loss levels.

Predicting churn has been a hotspot for academics for decades. Advances in ensembling and machine learning have allowed for better forecasts in this field. The work of Keaveney (Keaveney, 1995) offered the first study to address the importance of churn prediction. The model's survey found that Service Encounter Failure (SEF) and Core Service Failure (CSF) were the main reasons for customer attrition. The study of Padmanabhan (Padmanabhan *et al*., 2011) found that high-quality service and responsiveness from customers are the key indicators of turnover. The definition of models that carry out churn prediction is based on this research. One of the primary areas that has allowed churn prediction to be possible is machine learning. Models developed by Eiben (Eiben *et al*., 1998) that employ logistic regression for prediction, models developed by Athanassopoulos (Athanassopoulos, 2000), Kumar (Kumar *et al*., 2008), and Hoppner(Hoppner *et al*., 2018) that use decision trees for churn prediction, and models developed by (Amin and Maheshwari,2018) that use firefly based churn prediction. (Lariviere *et al*., 2005) Employed a random Forest-based model for churn prediction, whereas (Lemmens *et al*.,2006) used an ensemble model. In addition to the data certainty-based model put forth by (Coussement *et al*., 2006) presented a subscription-based churn prediction method. Kumar *et al*. [10] suggested a methodology for predicting customer attrition in financial institutions. (Vafeiadis *et al*., 2015) Suggested a comparative analysis of churn prediction machine learning algorithms. To further enhance the classifier's performance, this model incorporates ensemble-based approaches such as AdaBoost. Not a single one of those models addresses problems caused by imbalanced data, the existence of noisy and borderline occurrences, etc., even if the majority of them produce useful findings. Predictions were discovered to be significantly impacted by these factors. Nevertheless, it was not taken into account that analyzing consumer behavior is necessary to address these concerns. Additionally, attrition prediction algorithms are often influenced by company objectives. The health of the models cannot be determined just by classifier metrics as they do not account for costs. What follows is a discussion of a few methods that take the commercial perspective into account as well. A key factor in establishing the prediction cost is the customer lifetime value (CLV) (Bogaert and Delaere, 2023). Models also took into account a number of other derived variables, such as savings and profit (Bhattacharya, 1998). A churn prediction model and profit-driven models and metrics were developed by (Farris *et al*.,2010), while (Lalwani *et al*., 2021) developed a model that deals with costs.

(Xiao *et al*., 2016) put out a model for churn prediction that relies on choosing the right classifiers based on the data attributes. To build a layer-based model with iterative selection of a suitable model for the issue domain, this model provides an ensemble model based on group method data handling (GMDH) multiple classifiers. It is possible to make accurate predictions using this model because of its high type 1 and type 2 accuracy levels. Because it is an operational procedure, the suggested design is quite computationally intensive. (He and Ding, 2024) presented a churn prediction model in that relies on bagging and boosting. To enhance the prediction levels, it adopts a corrective technique and builds on top of bagged Decision Trees, AdaBoost, and Stochastic Gradient Boosting. The prediction process was shown to be significantly improved by the corrective mechanism despite its basic nature. In (Mienye and Sun, 2021) suggested a churn prediction model that is based on the Generalised Additive Model (GAM). The primary focus of this approach is to reconcile performance in order to produce results that can be understood, allowing the solution to directly align with the business objectives. In (Mirabdolbaghi *et al*.,2021) presented a comparable churn prediction model for the gaming sector. It has been noted that Random Forest, being the most popular and oldest ensemble, is utilized in many forms for churn prediction. (Lariviere and Vandenpoel, 2005; Saha *et al*., 2023) both present models that are based on random forests.

Methodology

Customer decision-making regarding the organization's service may be better understood with the help of churn prediction. An examination of the customer's actions and priorities in the past is necessary for this. Afterwards, it's necessary to examine how customers act. The use of machine learning models is critical for the detection of such trends. It is computationally difficult to analyze massive volumes of consumer behavior data in order to create accurate forecasts. There is a huge disparity in the costs associated with incorrect forecasts; for example, false negative predictions are 500 to 1000 times more expensive than false positive predictions. Due to the quick rate of change, the telecom industry is thought of as an extremely unstable setting. Therefore, accurate projections need ongoing assessments (Abdulsalam *et al*.,2022). Problems with data imbalance, data ambiguity, and unknown distribution make it hard to create such models. In order to generate cost-effective forecasts, a churn prediction model needs to take all of these factors into account. For the purpose of churn prediction, this work suggests a two-level stacked heterogeneous ensemble. Gradient boosted trees, and naïve bayes make up the initial stage of the suggested architecture, which is the creation of the heterogeneous ensemble. A combiner built using a one-class support vector machine (SVM) for minority upliftment makes up the second level of the proposed design.

Analysis of Data Inequality in Telecom Churn Records

A comprehensive examination of churn data (UCI Churn Data) reveals that it is extremely unbalanced. One argument in favour of this is the idea that there would be relatively few churners in any given organisation. The disparity between the percentage of those who churn and those who do not is seen in Figure 1. It is clear that churners make up 14% of the data instances, whereas non-churners account for 86% of the instances. A level of 5.9 was determined to be the imbalance level of UCI Churn Data.

Biased learning in classifiers is caused by data imbalance. Classifiers work on the assumption that the data is balanced. Classifier reliability suffers as a result of over-training the majority class and under-training the minority class caused by the abundance of available majority classes. Classifiers trained on unequal data should, therefore, provide different degrees of granularity for the majority and minority groups. By implementing such a strategy into classifiers,

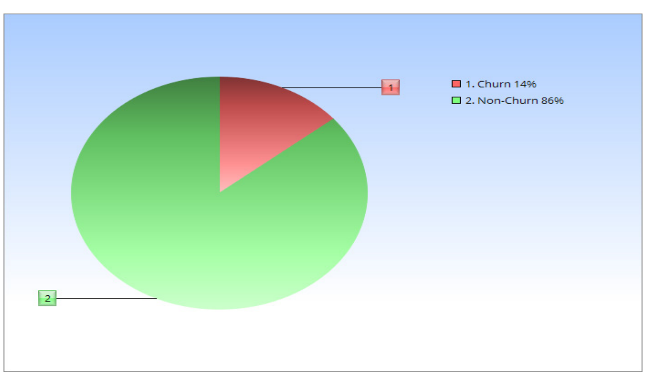

Figure 1:. Imbalance level of UCI churn data (in %)

their re-usability is reduced as they are fine-tuned to the imbalance levels in the training data. Ensembles 1 offer a more refined and doable answer to this problem.

Ensemble Techniques: A Condensed Overview

The goal of ensemble learning is to build a model with superior predictive performance than any of its individual learners by combining various learning algorithms, often known as base learners. Bagging, stacking, and boosting are three examples of frequent ensemble kinds. With the exception of boosting, which employs a single learner, all of these ensembles use several learners to forecast the training data, and the final forecast is generated by combining the prediction rules acquired from the various learners. There are a number of homogeneous models available in the literature, and the complexity of the rule aggregation phase typically discourages the formation of heterogeneous ensembles. The needs of the training data dictate the optimal ensemble to utilize, and each ensemble has its advantages and disadvantages.

By applying several base learners to the training data, stacking increases the likelihood of accurate predictions. Then, a heuristic combiner takes all of the base learners' outputs and uses them to produce an aggregated forecast. Experts in the field often provide such techniques for combining separate outcomes. This means that their performance may be fine-tuned through customization. Heterogeneity in the base learners allows effective management of many difficulties (such as imbalance, noise, borderline data, etc.), which is not achievable with homogeneous learners, even though homogeneous ensembles boost performance levels.

Base Learner Analysis

The proposed MMTCP model uses Naïve Bayes and Gradient Boosted Trees as its foundational learners. As a probability-based prediction method, Naïve Bayes maintains consistent performance levels even when data is imbalanced, or samples are on the edge of being considered valid. The foundation of Naïve Bayes, the Bayes model, is the implicit assumption of attribute independence, which leads to stable performance. The assumption of attribute independence states that each dataset attribute relies solely on the class attribute and has no other dependencies. Research shows that Naïve Bayes performs well even on data with low to moderate attribute dependencies, even though this assumption isn't totally accurate for churn prediction data. Naïve Bayes showed modest tolerance to unbalanced data because to its independent and probabilistic working nature. Since Naïve Bayes relies solely on probability calculations for its predictions, its computational complexity was shown to be O(n). As its foundational learner, gradient boosted trees (GBT) employ decision trees in its boosting-based ensemble.

Boosting uses a single decision tree instance as its foundational learner. Iteratively improving the solution involves tracking the levels of error at each stage and modifying the weights accordingly to get better solutions in the following stage. By training all classes equally, regardless of their representation in the training data, the training bias may be addressed by repeated lowering of error levels and weight adjustments. But noisy data, an inevitable part of churn data, has a significant impact on them. Because the boosting model's weights get misaligned when dealing with noisy data, inaccurate predictions are the result. Using m characteristics, n instances, and p iterations needed for GBT to converge, the computational complexity of Gradient Boosted Trees using decision trees as base learners is O (pmn log n).

Proposed Method (MMTCP) Multi-Model Telecom Churn Prediction

There are two tiers to the planned MMTCP. Training the models (Naïve Bayes and GBT), acquiring the predictions for the first level, and separating the suitable inputs for the second level are the objectives of the first level of operation. In order to enhance churn prediction, the second-level combiner employs minority upliftment with one-class support vector machines and final prediction aggregation. Implementations are carried out on Apache Spark using PySpark, and the whole procedure is parallelized for real-time performance.

Applying gradient-boosted trees (GBT) and naïve bayes (NB) on training data yields the trained model in the early stage of MMTCP. Consider two models, M_1 and M_2, and some training data, T.

T={(x1,C),(x2,C),…(xn,C)}, the class label C ∈{0,1} and the instances x_1, $x_2 \ldots x_n$ are defined.

Applying gradient boosted trees (GBT) and Naïve bayes (NB) on training data yields the trained model in the early stage of MMTCP. Consider two models, M_1 and M_2, and some training data, T.

. $T=\{(x_1,C),(x_2,C),\ldots(x_n,C)\}$, where $x_1, x_2 \ldots x_n$ are the instances and C is the class label where $C \in \{0,1\}$.

The training data *T* is passed to *NB* and *GBT* and the trained models T_{NB} and T_{GBT} are obtained.

$$T_{NB} = NB(T) \tag{1}$$

$$T_{GBT} = GBT(T) \tag{2}$$

Let the data to be predicted $P=\{x_{p1}, x_{p2} \ldots x_{pn}\}$ is evaluated by the trained models T_{NB} and T_{GBT} and the first level predictions of Naïve Bayes is given by,

$$Pred_{NB} = \left\{\left(x_{pi}, C_{NBi}\right) | x_{pi} \in P \wedge C_{NBi} \in \{0,1\}\right\} \quad \forall 1 \le i \le |P| \tag{3}$$

$$C_{NBi} = \begin{cases} 1 & if\ T_{NB}\left(x_{pi}\right) = 1 \\ 0 & Otherwise \end{cases} \quad \forall 1 \le i \le |P| \tag{4}$$

Similarly, the predictions for gradient boosted trees are given by,

$$Pred_{GBT} = \left\{\left(x_{pi}, C_{GBTi}\right) | x_{pi} \in P \wedge C_{GBTi} \in \{0,1\}\right\} \quad \forall 1 \le i \le |P| \tag{5}$$

$$C_{GBTi} = \begin{cases} 1 & if\ T_{GBT}\left(x_{pi}\right) = 1 \\ 0 & Otherwise \end{cases} \quad \forall 1 \le i \le |P| \tag{6}$$

Next, the predictions are divided into classes that are positive and negative. The list of churners predicted by GBT is called the positive class, whereas the list of non-churners projected by NB is called the negative class.

The positive and negative data segregations are given by

$$Pos = \left\{\left(x_i, C_i\right) | x_i \in Pred_{GBT} \wedge C_i = 1\right\} \quad \forall 1 \le i \le \left|Pred_{GBT}\right| \tag{7}$$

$$Neg = \left\{\left(x_i, C_i\right) | x_i \in Pred_{NB} \wedge C_i = 0\right\} \quad \forall 1 \le i \le \left|Pred_{NB}\right| \tag{8}$$

The list of churners from the prediction set P is contained in the positive data segregation (Pos), while the list of non-churners is contained in the negative data segregation (Neg). On the other hand, Pos and Neg's data are subsets of P. There are certain examples in P that Pos and Neg do not cover, and there are some examples in both Pos and

Neg.

A review of the findings indicates that

$$P \ne \{Pos \cup Neg\} \tag{9}$$

$$Pos \cap Neg \ne \varnothing \tag{10}$$

This is because GBT and NB showed different forecasts, and the data was imbalanced. The prediction set P still contains a number of common and undiscovered events. Therefore, in order to determine which P variables are churners and which are non-churners, a secondary data analysis is required. Phase one of minority upliftment is a secondary analysis, which examines data with inconsistencies and yields final conclusions. This necessitates sorting out the P data that doesn't fit into the Pos or Neg categories, as well as the data that does fit into both categories.

Data that has both positive and negative representations is recognized by

$$Common = \left\{\left(x_i, C_i\right) | (x_i, 1) \in Pred_{GBT} \wedge (x_i, 0) \in Pred_{NB}\right\} \quad \forall 1 \le i \le \left|Pred_{GBT}\right| \tag{11}$$

Unrepresented data is identified by finding the negative representations in $Pred_{GBT}$ and positive representations in $Pred_{NB}$. This is given by,

$$AdditionalN = \left\{ \left(x_i, C_i \right) | (x_i, 0) \in Pred_{GBT} \right\} \tag{12}$$

$$\forall 1 \leq i \leq \left| Pred_{GBT} \right|$$

$$AdditionalP = \left\{ \left(x_i, C_i \right) | (x_i, 1) \in Pred_{NB} \right\} \tag{13}$$

$$\forall 1 \leq i \leq \left| Pred_{NB} \right|$$

The input data for the minority upliftment model is obtained by considering all the data from *Common, AdditionalN* and *AdditionalP* and is given by,

$$Data_{L2} = Common \cup AdditionalN \cup AdditionalP \tag{14}$$

Hence the final prediction set for positive and negative classes is given by

$$FinalPos = \left\{ \left(x_i, C_i \right) | \begin{matrix} (x_i, 1) \in Pos \wedge (x_i, 1) \notin \\ Common \wedge (x_i, 0) \notin Common \end{matrix} \right\} \tag{15}$$

$$\forall 1 \leq i \leq \left| Pos \right|$$

$$FinalNeg = \left\{ \left(x_i, C_i \right) | \begin{matrix} (x_i, 0) \in Neg \wedge (x_i, 1) \\ \notin Common \wedge (x_i, 0) \notin Common \end{matrix} \right\} \tag{16}$$

$$\forall 1 \leq i \leq \left| Neg \right|$$

The second-level combiner is designed to support minority upliftment using one-class support vector machines. It has been noted that classifiers are negatively impacted by data that is unbalanced. There are many more examples of the majority class than the minority classes. Therefore, when a classifier is trained with this data, the low minority class levels lead to under-training, and the vast number of majority classes tend to over-train. The forecasting of minority classes (churners) is problematic, in contrast to the often accurate forecasting of majority classes (non-churners)..

One Class SVM

By only training Support Vector Machines (SVMs) with examples belonging to one class, the One-Class SVM method achieves binary classification. It uses the training data to generate a hyperplane. Each instance outside the hyperplane represents the alternative class, while every instance inside the hyperplane represents the training class. One-class support vector machines (SVMs) are a subset of classifiers designed to handle data with imbalances. The suggested design uses it as a secondary-level booster instead of a mainstream classifier because of its poor performance on churn data. In order to train the one-class SVM model, the instances representing churn, which are considered minority classes, are separated. When it comes to making predictions, one-class support vector machines (SVMs) are like solving a dual quadratic programming problem; they are provided by

$$T_{SVM} = \min_{\alpha} \frac{1}{2} \sum_{ij} \alpha_i \alpha_j K\left(x_i, x_j \right) \tag{17}$$

Subject to the constraints

$$0 \leq \alpha_i \leq \frac{1}{vl}$$

and

$$\sum_i \alpha_i = 1$$

Where l is the number of points in the training dataset, α_i is a Lagrange multiplier, v is the control parameter that balances the hyperplane's distance from the origin, and the number of data points it contains, and $K(x_i,x_j)$ is the kernel function.

In this case, ϕ translates training vectors to a high dimensional feature space from input space X. Given that the suggested one-class SVM employs a linear kernel, the kernel function may be defined as,

$$K\left(x, y \right) = \prod_{1}^{d} \phi\left(d \right) \tag{18}$$

where *d* is the number of dimensions in the training data.

Let T_{SVM} be the trained one-class SVM model trained using the minority class instances *Min,* which is given by,

$$Min = \left\{ \left(x_i, C_i \right) | C_i = 1 \right\} \qquad \forall 1 \leq i \leq \left| T \right| \tag{19}$$

$Data_{L2}$, obtained by aggregating the common and unrepresented data from *Pos* and *Neg* is passed to the one-class SVM, and the predictions are obtained as,

$$CommonPred = T_{SVM} \left(Data_{L2} \right) \tag{20}$$

To get the final forecasts, these are mixed with the final sets of positive and negative predictions.

$$FinalPred = FinalPos \cup FinalNeg \cup CommonPred \tag{21}$$

Due to data inconsistency and noise in the telecom churn statistics, a minority upliftment phase is being implemented. A study of data based on business goals shows that cost-based projections, as opposed to predictions based on traditional performance indicators, have a good connection with the aims of the company. Therefore, the suggested approach was evaluated based on costs.

Cost Based Model

The suggested model's efficacy is heavily dependent on its cost. It is unnecessary to take cost into account when dealing with data that has constant and balanced distributions [25]. Unfortunately, traditional classifier measures like accuracy won't cut it when dealing with data that is imbalanced and noisy, like churn data. In addition, the correlation between the cost components and traditional metrics is rather weak. As a result, there is little alignment with the objectives of the company. In addition, a number of indicators provide equal weight to all predictions. From a commercial standpoint, though, the costs associated with each forecast are unique. In the churn domain, there is a cost associated with correct classifications as well. When classifications are accurate, costs are lower; when they are inaccurate, costs are higher. An expense-conscious approach is thus necessary for the churn prediction domain.

Figure 2 displays the results of using cost estimates for churn prediction. It was noted that there were significant variances in the cost of forecasts. New customers are represented by customer inflow, whereas churners are represented by customer outflow.

Each prediction's associated cost is shown in an oval shape. There are two choices available to TPs who get their predictions right: remain or go. When they remain, the cost is equal to the offer cost plus administrative cost; when they depart, it's equal to the customer's lifetime value plus administrative cost. A client who is mistakenly believed to be a churner (FP) ends up staying with the company, costing the same as the offer cost (co) plus the administrative cost (ca). The cost associated with a churner who is incorrectly identified as a retained customer (FN) is equal to their Customer Lifetime Value (CLV). No expense will be incurred for a retained client (TN) that is accurately forecasted.

The cost matrix for the forecasts is shown in Table 1. This matches the confusion matrix exactly. This was taken from the Bahnsen *et al.* cost-sensitive churn prediction model [18].

Where CLV stands for customer lifetime value, co and ca stand for offer cost and administrative cost, and γ is the percentage of customers that are kept off the list of churners.

The estimates are used to compute the cost. The method for calculating costs is provided by,

$$Cost_i = y_i\left(c_i C_{TP_i} + \left(1-c_i\right)C_{FN_i}\right) + \left(1-y_i\right)\left(c_i C_{FP_i} + \left(1-c_i\right)C_{TN_i}\right) \tag{22}$$

....

where c_i and y_i are the predicted and the actual class labels, $y \in \{0,1\}, c \in \{0,1\}$ refer to non-churn and churn, respectively, CTP, CFP, CTN and CFN are obtained from the cost matrix.

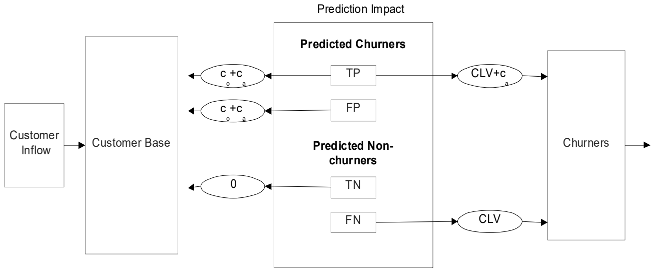

Figure 2: Financial impact of churn predictions

Table 1: Cost matrix

Actual / *Predicted*	*Churn*	*Not churn*
Churn	γ (co + ca) +(1-γ)(CLV + ca)	co + ca

Customer Lifetime Value (CLV), offer cost (co), and administrative cost (ca) are the three variables that make up the cost equation. One way to determine CLV is to use the customer turnover probability as a starting point for estimating the customer's future value. Three criteria can be used to determine the CLV, according to (Saha *et al.*, 2023). It was determined that CLV >> co+ca when comparing the cost of the customer's CLV to the aggregated cost of the offer and administrative expenses. Additionally, offering discounts and deals to clients is typically seen as an integral aspect of a telecom company's overall business plan. Offering more to customers usually makes them feel better about the company, even when it costs money. It may even delay their choice to churn despite being intangible. Therefore, providing clients with more offerings is an investment, not a loss. Therefore, this study proposes a loss function that addresses real loss issues instead of the conventional cost factor. Here is the loss function that has been suggested:

$$\text{Total Loss} = \text{Loss}_{\text{UA}} + \text{Loss}_{\text{Pred}} \tag{23}$$

Whereas $\text{Loss}_{\text{Pred}}$ represents a loss that happened because of an incorrect forecast and Loss_{UA} represents an inevitable loss. Customer turnover that persists despite promotional offerings is an inevitable loss. This value is corresponding to a percentage of TPs. It's the False Negative predictions that correspond to $\text{Loss}_{\text{Pred}}$. Here are the real degrees of loss:

$$\text{Total Loss} = \text{TP}\left(1-\right)\left(\text{CLV} + \text{c}_{\text{a}}\right) + \left(\text{FN} \times \text{CLV}\right) \tag{24}$$

Where *γ* represents the level of retained customers, ca correspond to the administrative cost and CLV represents the Customer Lifetime Value.

According to the suggested loss function, the most important parts are those that deal with the customer's lifetime value (CLV). Within the confusion matrix, these components stand for the levels of True Positive (TP) and False Negative (FN). Therefore, the first objective for a churn prediction classifier should be to provide low FN levels and high TP levels.

To give proper weight to minority class suitable TP and FN value predictions, the suggested design uses a secondary level minority upliftment combiner with One-Class SVM. In addition to improved prediction rates, the minority upliftment combiner that is suggested using One-Class SVM shows remarkable resistance to imbalance and noise. By limiting training to instances from the minority class, under-training owing to imbalance is rendered obsolete in the context of One-Class SVM. Reducing the quantity of training data, which in turn lowers training time and computational cost, is an additional benefit of employing cases from minority classes for training

Results and Discussions

Both the suggested loss function and more traditional performance criteria are used to evaluate the MMTCP model's efficacy. The GBT, Naïve Bayes, and One-Class SVM, which are the separate parts of the proposed MMTCP, and the methods for integrating them, are built in Apache Spark with PySpark as the language of choice. Utilising UCI churn data for prediction, the outcomes are evaluated using both conventional metrics for classifier performance and the suggested loss function. We check our results against those of previously published classifiers and churn prediction methods. In Table 2, you can see the characteristics of the UCI churn statistics.

Analysis of Performance Metrics (MMTCP)

Table 3 displays the performance level that the suggested MMTCP exhibits.

With a true positive rate of 0.892, we can see that 89% of the churners were accurately recognized in terms of churn prediction levels. With a false-negative rate of only 0.107, the suggested MMTCP model has lost 10% of its churners because it wrongly identified them as non-churners. Moreover, the suggested MMTCP model shows effective results with a recall of 89%, an accuracy of 28.7%, and an AUC of 0.77. Due to imbalance, the first-level classifiers, GBT and NB, are affected; nevertheless, the upliftment phase helps to significantly reduce this influence, and the outcomes are successful. One-class SVM's minority upliftment is responsible for the lower FNR levels and higher TPR levels.

Comparison based on Standard Classifier Models

Table 4 shows the prediction levels after comparing MMTCP results with those of other classifier models, such as Decision Tree, SVM, and ANN.

The churn prediction rate, hit rate, and lift coefficient are the metrics used for comparison. A higher number indicates a more favorable outcome. Overall, the suggested MMTCP model has the best prediction levels across the board.

Table 2: Properties of UCI churn data

Data segregation level	*3333 (Training Data), 1667 (Test Data)*
Attributes	19 Attributes + 1 Class
Total non-churn	2850
Total churn	483
Attribute type	16 Numeric+ 4 Nominal
Imbalance ratio (Non-Churn : Churn)	6:1

Table 3: Performance metrics (MMTCP)

Metric	*Value*
TPR	0.892857143
FNR	0.107142857
Recall	0.892857143
AUC	0.287769784
Precision	0.7749109
FPR	0.343035343

Table 4: Comparison of classifier models

Technique	*Churn detection rate*	*Hit rate*	*Lift coefficient*
Proposed MMTCP	0.8929	0.2878	2.6401
SVM	0.6546	0.2632	2.312
ANN	0.4686	0.2372	1.643
Decision Tree	0.6103	0.2265	1.725

Figure 3 displays the churn detection rate for MMTCP, SVM, ANN, and Decision Tree. The churn detection levels of MMTCP were found to be 24 to 43% higher than those of the previous models. Due to its reliance on data levels (the amount of representations of churn and non-churn entries) during training, ANN, SVM, and the entropy-based divisions of decision trees are very susceptible to imbalance, which in turn leads to worse predictions. At the first level, superior detection rates are provided by probability-based predictions and the conditional independence of Naïve Bayes, which demonstrate immunity towards imbalance. Boosting also assists in effective error management. The minority upliftment utilizing One-Class SVM, however, provides extra boosting.

Comparison with Existing Churn Prediction Techniques

A comparison of churn prediction rates with existing churn prediction techniques is performed and the results are tabulated in Table 5.

Comparisons are performed in terms of churn detection rates and false negative prediction levels. The proposed MMTCP model exhibits the best churn detection rate.

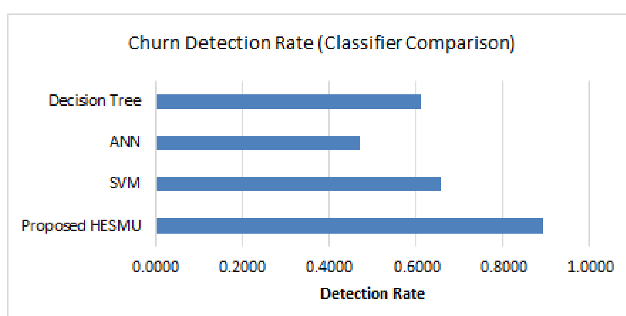

Figure 3. Churn detection rate

Table 5: Performance comparisons with existing churn prediction models

Technique	Churn detection rate	False negative rate
Proposed MMTCP	0.8929	0.1071
GMDH (Lemmens, 2006)	0.8803	0.1197
Bagging (Bogaert, and Delaere, (2023)	0.8462	0.1538
Boosting (He and Ding, 2024)	0.8281	0.1719
Generalized Additive Models (Coussement, and VanDen Poel,2006)	0.8404	0.1596
Ensemble Churn Prediction (Xiao et al., 2016)	0.8353	0.1647
Random Forest (Saha et al., 2023)	0.8492	0.1508
Balanced Random Forest (Saha et al., 2023)	0.8532	0.1468

The contrast is illustrated graphically in Figure 4. Compared to churn prediction models found in the literature, the suggested MMTCP model has a 1 to 7% greater churn detection rate. In a similar vein, various models display false negative rates ranging from 11 to 17%, while the suggested MMTCP exhibited the lowest at 10%.

Comparison based on loss

Table 6 compares the outcomes of current classifier models according to the prediction loss levels.

The suggested loss function shows the overall loss as a sum of both the inevitable loss and the loss caused by incorrect forecasts. As the term implies, unavoidable loss is impossible to escape. Even though the company offered them promotions, they still decided to quit, making them a small proportion of the churners that were accurately anticipated. To achieve their overarching goal of reducing overall loss, churn prediction models typically maximize the inevitable loss while minimizing the damage caused by incorrect predictions.

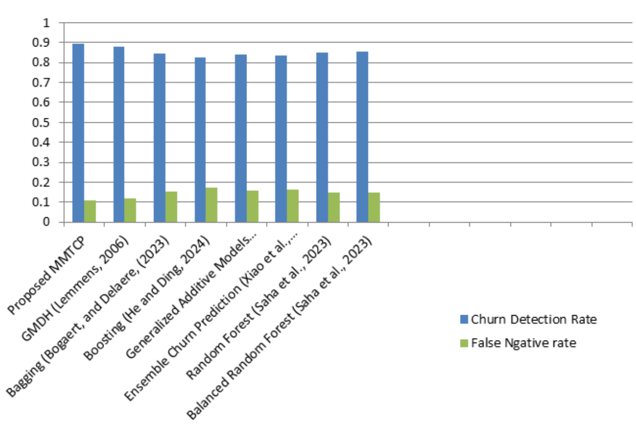

Figure 4: Churn detection rate

Table 6. Comparison based on loss

Technique	Total loss detection rate	Unavoidable loss	Loss due to wrong prediction negative rate
Proposed MMTCP	16860.0	12060.0	4800.0
GBT	22308.3	9708.3	12600
Naïve Bayes	24403.8	8803.8	15600
Decision tree	24543.5	8743.5	15800
Random forest	29712.4	6512.4	23200

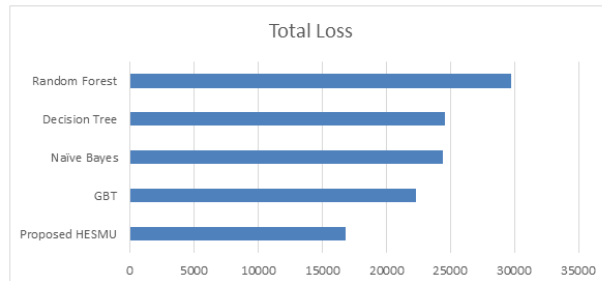

Figure 5. Total loss

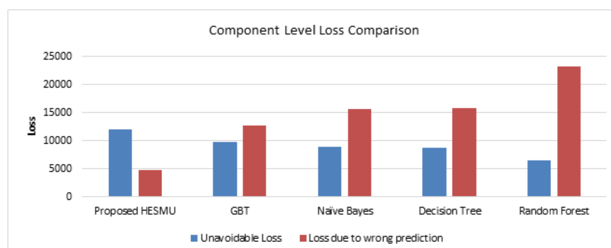

Figure 6: Component level loss comparison

Figure 5 displays the total loss that the models that are being compared have experienced. The suggested MMTCP model shows the least overall loss when compared to the other models, with loss levels reduced by 1.3 to 1.7 times.

Figure 6 shows a breakdown of the overall loss, broken down into loss that could not be avoided and loss that was caused by incorrect projections. A churn prediction model's primary objective is to maximize the inevitable loss while minimizing the loss caused by incorrect forecasts. The inevitable loss in MMTCP was 1.2 to 1.8 times larger, whereas the loss owing to incorrect forecasts was 2.6 to 4.8 times lower.

Conclusion

Having the correct metrics for prediction analysis that align with the organization's business goals is essential in the everchanging telecom industry. An improved churn prediction model using heterogeneous ensemble stacking and minority upliftment is presented in this research. The use of diverse classifiers results in the formation of an ensemble based on stacking. Inherent problems, such as imbalance, noise, and borderline data, might impact algorithmic forecasts. By establishing a heterogeneous framework, the suggested MMTCP method gets over these problems. A two-tiered processing design has been suggested. Gradient boosted trees and naïve bayes make up the first level, which aims to enhance prediction efficiency and stability. To help minority upliftment, the initial level predictions are sent to the One-Class SVM. A loss function is suggested that divides the loss into two parts: the part that cannot be avoided and the part that is caused by incorrect prediction. Compared to conventional classifier models, experimental results show churn detection rates that are 24 to 43% higher. The churn prediction values are 1 to 7% higher when compared to models found in the literature. The MMTCP model outperforms state-of-the-art classifier methods in terms of overall loss, which is 1.3 to 1.7 times lower. The goal of future research is to develop a cost-sensitive model that can aggregate or create domain-based features to further minimize loss and the number of false positives.

Acknowledgments

The authors extend their thanks and gratitude to the journal editors and reviewers for their valuable comments.

References

Abdulsalam, S. O., Arowolo, M. O., Saheed, Y. K., and Afolayan, J. O. (2022). Customer churn prediction in telecommunication industry using classification and regression trees and artificial neural network algorithms. *Indonesian Journal of Electrical Engineering and Informatics (IJEEI)*, *10*(2).

Ahmed, A. A., Maheswari, D. (2017). Churn prediction on huge telecom data using hybrid firefly based classification. Egyptian Informatics Journal, 18(3), 215–220.

Amin, A., Al-Obeidat, F., Shah, B., Adnan, A., Loo, J., Anwar, S. (2018). Customer churn prediction in telecommunication industry using data certainty. Journal of Business Research, 94, 290–301.

Amin, A., Anwar, S., Adnan, A., Nawaz, M., Alawfi, K., Hussain, A., Huang, K. (2016). Customer churn prediction in the telecommunication sector using a rough set approach. Neurocomputing, 237, 242–254.

Amin, A., Shah, B., Khattak, A. M., Moreira, F. J. L., Ali, G., Rocha, A., Anwar, S. (2018). Cross-company customer churn prediction in telecommunication: A comparison of data transformation methods. International Journal of Information Management, 46, 304–319.

Athanassopoulos, A. D. (2000). Customer satisfaction cues to support market segmentation and explain switching behavior. Journal of Business Research, 47(3), 191–207.

Bhattacharya, C. B. (1998). When customers are members: Customer retention in paid membership contexts. Journal of the Academy of Marketing Science, 26(1), 31–44.

Bogaert, M., and Delaere, L. (2023). Ensemble Methods in Customer Churn Prediction: A Comparative Analysis of the State-of-the-Art. Mathematics, 11(5), 1137.

Coussement, K., and Van Den Poel, D. (2006). Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. Expert Systems With Applications, 34(1), 313–327.

Eiben, A. E., Koudijs, A. E., and Slisser, F. (1998). Genetic modelling of customer retention. In Lecture notes in computer science (pp. 178–186).

Farris, P. W., Bendle, N. T., Pfeifer, P. E., and Reibstein, D. J. (2010). Marketing metrics: the definitive guide to measuring marketing performance. Choice Reviews Online, 48(01), 48–0373.

He, C., and Ding, C. H. Q. (2024). A novel classification algorithm for customer churn prediction based on hybrid Ensemble-Fusion model. Scientific Reports, 14(1).

Höppner, S., Stripling, E., Baesens, B., Vanden Broucke, S., and Verdonck, T. (2018). Profit driven decision trees for churn prediction. European Journal of Operational Research, 284(3), 920–933.

Keaveney, S. M. (1995). Customer switching behavior in service Industries: An exploratory study. Journal of Marketing, 59(2), 71.

Kumar, D. A., and Ravi, V. (2008). Predicting credit card customer churn in banks using data mining. International Journal of Data Analysis Techniques and Strategies, 1(1), 4.

Lalwani, P., Mishra, M. K., Chadha, J. S., & Sethi, P. (2021). Customer churn prediction system: a machine learning approach. Computing, 104(2), 271–294

Lariviere, B., and Vandenpoel, D. (2005). Predicting customer retention and profitability by using random forests and regression forests techniques. Expert Systems With Applications, 29(2), 472–484.

Lemmens, A., and Croux, C. (2006). Bagging and boosting classification trees to predict churn. Journal of Marketing Research, 43(2), 276–286.

Mienye, I. D., and Sun, Y. (2021). Performance analysis of cost-sensitive learning methods with application to imbalanced medical data. Informatics in Medicine Unlocked, 25, 100690.

Mirabdolbaghi, S. M. S., and Amiri, B. (2022). Model Optimization Analysis of Customer Churn Prediction Using Machine Learning Algorithms with Focus on Feature Reductions. Discrete Dynamics in Nature and Society, 2022, 1–20.

Padmanabhan, B., Hevner, A., Cuenco, M., and Shi, C. (2011). From information to operations. ACM Transactions on Management Information Systems, 2(4), 1–21.

Saha, L., Tripathy, H. K., Gaber, T., El-Gohary, H., and El-Kenawy, E. M. (2023). Deep churn prediction method for telecommunication industry. Sustainability, 15(5), 4543.

Vafeiadis, T., Diamantaras, K., Sarigiannidis, G., and Chatzisavvas, K. (2015). A comparison of machine learning techniques for customer churn prediction. Simulation Modelling Practice and Theory, 55, 1–9.

Xiao, J., Jiang, X., He, C., and Teng, G. (2016). Churn prediction in customer relationship management via GMDH-Based Multiple Classifiers ensemble. IEEE Intelligent Systems, 31(2), 37–44.