ISSN: 2229-6956 (ONLINE)
DOI: 10.21917/ijsc.2021.0353

MINIMAX PROBABILITY-BASED CHURN PREDICTION FOR PROFIT MAXIMIZATION

V. Jude Nirmal

Department of Computer Science, Bharathidasan University, India

Abstract

Churn prediction has become a significant requirement for all customer centric organizations. Accurate prediction of churn can effectively improve customer loyalty and improve profits for the organization. This work presents an effective model that uses a combination of ensemble learning and minimax probability machines to provide a churn prediction system. The model has its major focus towards improving the profitability of the organization. The ensemble learning model has been designed to be computationally efficient, while the weight factors used in the minimax probability machines ensures reduction in losses, hence ensuring profitability. Experiments were performed and comparisons with existing models indicates that the model shows high performance, with 8% improved accuracy levels, indicating improved churn predictions.

Keywords:

Churn prediction, Ensemble learning, Minimax Probability Machines, Extra Trees Classifier, Profit Maximization

1. INTRODUCTION

Current technology induced world has brought customers much closer to the organizations. This has also made every customer-based industry experience the issue of churn [1]. Customer churn is the process of a customer leaving an organization to opt for another competitor organization for various reasons. Churn has become an unavoidable component in several domains like IT services, social networking services, internet service providers and mobile services [2] [3]. Customer Relationship Management (CRM) has become a vital part of every organization. Effectively utilizing CRM in an organization can result in improving customer loyalty to a large extent, hence better customer retention. Although this process is necessary for every organization, the main requirement for churn reduction has raised in the telecommunication industry.

Rapid changes in the players, many available players, frequent introduction of new technical features and attractive plans has provided ample opportunities for customers to shift between operators [4]. Ease of shift and low cost of shifts has also played major roles in the large number of shifts occurring in the telecom environment [5]. According to management principles, it was observed that it costs higher to gain a customer compared to retaining an existing customer. Several organizations lose customers mainly due to ineffective churn prediction models or the absence of such models in their CRM architecture. Hence, maintaining an effective churn prediction system is of vital importance in any customer centric organization.

Churn prediction has become an interest for research community because of the challenges involved in correctly determining churn. Initial contributions in this domain included machine learning models [6]. These were quickly followed by ensemble models [7], metaheuristic models [8] and neural network models [9].

However, there still exists scope in the domain due to the large number of intrinsic challenges exhibited in the domain, such as data imbalance and noise [10]. Issues due to imbalance occurs due to the unbalanced nature of the occurrence of churners and non-churners. Due to the real-time nature of the domain, the number of non-churners is usually much higher than the number of churners. Further, some probable non-churners also have the possibility of moving out of the organization due to personal constraints. Such records should be identified and eliminated, as they act as noise tend to corrupt the decision rules.

This work proposes an effective model that can operate on telecom churn data to handle the above-mentioned issues to provide an effective and profit oriented churn prediction model.

2. RELATED WORK

Determining churners accurately has become a major requirement of several customer-based industries. In specific, telecom industry exhibits the maximum requirement due to the large number of competitors and the ease at which customers can switch between operators. This section presents recent contributions in the churn prediction domain.

A churn prediction model that uses a combination of logistic regression and logit boost models for the prediction process was proposed by Jain et al. [11]. The models have been tested using the WEKA Machine learning tool and analysis of performance has been done using the dataset obtained from the American telecom operator Orange. A Latent Dirichlet Allocation (LDA)-based model for predicting customer churn was presented by Slof et al. [12]. This work uses a duration-based model for predicting the propensity of a customer to churn. The work also employs a risk-based model that also determines the reason for the risks. Call logs between the customer and the call center executive are used by the LDA model for analysis. Other similar duration-based models include works by Lariviere et al. [13] and Jamal et al. [14]. A similarity forests-based model for churn prediction was proposed by Infante et al. [15]. This technique determines the relevance of each existing customer-based on social network analysis models and a combination of binary classification techniques. This is a graph-based model that extracts the behavior patterns of churners and non-churners.

Neural networks are one of the most preferred models for the churn prediction process. A model that is-based on neural networks for identifying churners was proposed by Omar et al. [16]. This work uses the basic multilayer perceptron architecture for the prediction process. Vector embedding-based model that uses deep learning models for churn prediction was proposed by Cenggoro et al. [17]. This work is solely-based on creating a highly interpretable model, as deep learning models by nature exhibits less interpretability. The model generates vector for each customer instance, and the vectors were identified to be highly discriminative in nature, making the model highly interpretable and explainable. The model also deals with providing choices that can be used to convert susceptible churners into non-churners. Behavior-based deep learning model for churn prediction was proposed by Alboukaey et al. [18]. It is-based on the fact that behavior of customers varies from time to time. Hence it is mandatory to watch their behavior continuously to effectively determine their tendency to churn. Time series-based modelling has been implemented and CNN-based model is used for the prediction. Several similar models have been designed to perform temporal predictions. Such models include hourly-based analysis by Zaratiegui et al. [19], daily analysis model by Wangperawong et al. [20] and weekly analysis model by Óskarsdóttir et al. [21].

An evolutionary model used for churn prediction was proposed by Wai-Ho et al. [22]. A graph theory-based model that utilizes social network analysis was proposed by Kostic et al. [23]. This uses graph-based modelling for the identification of churners. The model aims to identify significant nodes in the graph model that serves as the major components for the churn identification process. A profit-based churn prediction architecture was presented by Hoppner et al. [24]. This work mainly concentrates on creating a model that has its major focus on obtaining profitable predictions rather than accurate predictions. The model further concentrates on improving the interpretability of the model for better understanding of the predictions. Other similar profit driven models include works by Maldonado et al. [25], and Stripling et al. [26].

3. MINIMAX PROBABILITY-BASED CHURN PREDICTION (MPCP)

Predicting churn accurately is of significant importance, as it involves customers and is highly time constrained. A delay in the predictions or in the secondary actions is bound to result in the loss of customer. Losing a customer results in huge losses for the organization, hence it becomes mandatory to perform strategies that can aid in customer retention. Further, gaining new customers is considered to be much costlier compared to retaining existing customers. Hence it becomes mandatory to consider the loss factors prior to determining possible churners.

This work proposes an ensemble-based model that is integrated with minimax probability machines for predicting churners from telecom data. The Minimax Probability-based Churn Prediction (MPCP) model has been designed in three stages; the initial stage performs data preprocessing, the next stage uses extra trees classifier model to predict the probability levels of the customer being a churner, and the final phase integrates minimax probability to provide profit-based final predictions.

3.1 DATA ANALYSIS AND STANDARDIZATION

This phase has been designed to perform churn prediction on telecom data. Hence this phase exhibits significant importance, as telecom data is usually huge and contains several noise elements and missing information. The training data is explanatorily analyzed to understand the attributes. Attributes that are least significant are eliminated. These includes string data and identifying information. Both these types of data do not provide significant knowledge to the machine learning model. Hence, are eliminated. Further, users with no call logs are also eliminated, as they are considered to be new users. The next type of issue to be handled is categorical attributes. Such attributes are common in real-time data; however, they cannot be handled by machine learning models. Categorical attributes are converted into numerical values using one-hot-encoding techniques. Although this process tends to increase the size of the data to a large extent, it becomes necessary for effective functioning of the model.

3.2 EXTRA TREES (EXTREMELY RANDOMIZED TREES) CLASSIFIER-BASED PROBABILITY PREDICTION

The data is then segregated into two sections; training and test data. The data division is performed such that training data comprises of 80% of the data and the remaining 20% of the data is used for testing. An ensemble-based bagging model is used for the initial phase prediction process. This work is-based on Extra Trees classifier model, which is a bagging-based model. However, several key differences exist between general bagging models and the extra trees model. Bagging model generally uses bootstrapping for training data creation. In the extra trees model, bootstrapping is not applied. The entire training data is passed to all the base learners. Base learner is a tree-based model that is used for decision rule building. Variation in the rule building process is created by selecting the node split points in random, rather than using the entropy values to create attribute splits. The major advantage of extra trees model is that it is much faster than the traditional random forest model.

Input to the model is composed of churn data. The problem domain is considered as a binary classification problem. Data is passed to each base learner to determine the decision rules. The decision rules generated and are modified in this work to provide the probability of the prediction of each class, rather than the final prediction. The output from this phase contains a tuple containing two values; the probability values of a record being a non-churner and churner. Regular operations consider the class exhibiting the highest probability value as the final prediction. This work passes the probability values to the next phase to perform profit maximization-based prediction.

3.3 MINIMAX PROBABILITY-BASED FINAL PREDICTION FOR PROFIT MAXIMIZATION

This phase uses minimax probability-based operations to maximize profits in the final predictions. The proposed model is-based on the minimax probability machines by Lanckriet et al. [27]. Minimax probability machine is-based on creating a hyperplane between the two available classes, and also tends to minimize errors. The model aims to minimize the worst-case probability of misclassification. It considers data from two classes X_1 and X_2. The mean and co-variance of the classes are calculated and recorded to be μ_i and σ_i respectively, where $i \in \{1, 2\}$. Hence, the separating hyperplane will be of the form $w^t x+b=0$, where w is the weights of classes, x is the actual data, and b is the bias constant. The maximum probability machine is considered as a chance constrained problem. Final predictions are obtained from the minimax probability machine-based on the weights for the churners and non-churners. The weights for churners is set to 0.7, and that of non-churners is set to 0.3. This is due to the fact that churners tend to create higher losses than non-churners, hence significance for identifying a churner is considered to be higher than the non-churner. The weights ensure that higher significance is provided to predicting a churner, hence resulting in higher profits in the prediction process. Final predictions are obtained by applying the probability values and weights to the equation. The class with highest value is considered as the final prediction.

4. RESULTS AND DISCUSSION

The churn prediction model MPCP has been implemented using Python and Scikit libraries. Analysis of the performance of MPCP is performed by applying it on the publicly available churn prediction data [28]. The dataset is composed of records obtained from 3,333 customers. The data is composed of 21 attributes. The class attribute is named as churn, and is a Boolean attribute. The data contains a mixture of categorical and numerical attributes. Details about the attributes are presented in Table.1.

Table.1. Attribute Description

Attribute	Data Type	Attribute	Data Type
State	Object	Total eve calls	Int64
Account length	Int64	Total eve charge	Float64
Area code	Int64	Total night minutes	Float64
Phone number	Object	Total night calls	Int64
International plan	Object	Total night charge	Float64
Voice mail plan	Object	Total intl minutes	Float64
Number vmail messages	Int64	Total intl calls	Int64
Total day minutes	Float64	Total intl charge	Float64
Total day calls	Int64	Customer service calls	Int64
Total day charge	Float64	Churn	Bool
Total eve minutes	Float64	-	-

An analysis of the ROC curve is presented in Fig.1. ROC curve is constructed with False Positive Rate (FPR) in x-axis and True Positive Rate (TPR) in y-axis. The graph represents the relationship between correct prediction of churners and false alarms. Low false alarms with high prediction levels represent good performance. The Fig.1 shows zero FPR levels, which shows that the model exhibits no false alarms. Similarly, the Fig.1 also shows high TPR levels indicating that the model exhibits effective identification of churners.

The PR plot representing the precision and recall values is shown in Fig.2. PR plot shows recall in the x-axis and precision in y-axis. Precision represents the level of actual churners in the identified list of churners, and recall represents the level of churners identified from the actual list of churners. Both precision and recall values are required to be high for a model to perform effectively. The Fig.2 shows excellent precision and high recall levels. This shows that the proposed model can effectively identify churners, and also that the identified churners are also highly precise with low errors.

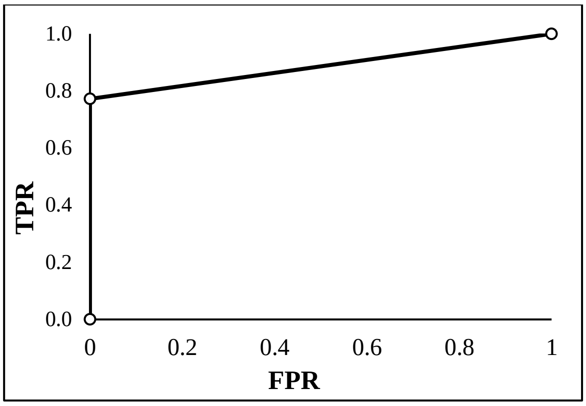

Fig.1. ROC Plot

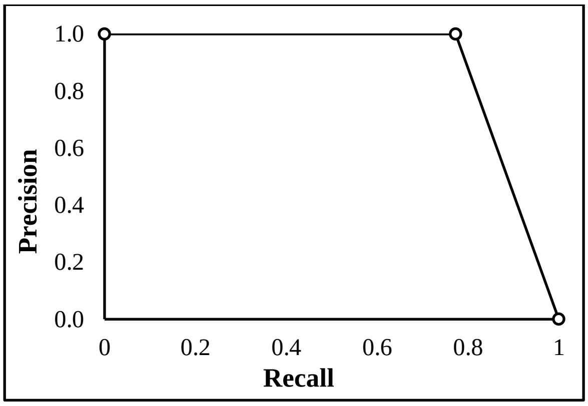

Fig.2. PR Plot

Performance of the MPCP model is shown-based on existing classifier performance metrics. The Table.2 could be analyzed to identify that the MPCP model depicts low FPR and FNR levels, and very high TPR and TNR levels. This shows the efficiency of the prediction model in effectively classifying the instances. Further, the aggregated performance metrics Accuracy and F1-score also shows high performance.

Table.2. Performance of MPCP Model

Metric	MPCP
FPR	0
TPR	0.773
TNR	1.000
FNR	0.227
Recall	0.773
Precision	1.000
Accuracy	0.970
F1-Score	0.872

A comparison of the performance of the MPCP model is performed with Deep Learning - Vector Embedding Model (DL-VEM) in [16]. Performance comparison is in terms of accuracy and F1-Score. Charts could be perceived to identify that the MPCP model depicts higher values for both accuracy and F1-Score. This indicates that the MPCP model exhibits effective identification of churners from the telecom data.

The performance comparison in tabulated form is presented in Table.3. The best performance is shown in bold. It could be observed that the MPCP model exhibits 8% improved accuracy levels and 6% improved F1-Score levels, indicating that the model is suiTable.to be used in real-time.

Table.3. Performance Comparison of MPCP

	MPCP	**DL-VEM**
Accuracy	0.97	0.898
F1-Score	0.87	0.811

5. CONCLUSION

Predicting churn effectively not only aids in customer retention, but also provides an effective way to improve the profits in the organization. This work presents a probability-based model MPCP, which uses an aggregation of machine learning model and minimax probability machines to provide profit-oriented churn predictions. Experiments were performed using the churn prediction data available from Kaggle. Comparisons were performed with existing models in literature. Comparisons-based on accuracy indicates that the MPCP model exhibits 8% improvements and F1-Score indicates 6% improvements. The upside of the MPCP model is that it is highly generic in nature. Even though analysis has been performed on telecom data, the model can be used on all types of data. The usage of weight factors in the architecture ensures that the model maximizes profit to a large extent. Although the model provides comparatively better performance, the model still has scope for improvement. Hence future extensions of the model will deal in improving the performance of the model.

REFERENCES

[1] A.M. Hughes, "Churn Reduction in the Telecom Industry", Available at http://www.dbmarketing.com/telecom/churnreduction.html, Accessed at 2020.

[2] C. Gronroos, "A Service Quality Model and its Marketing Implications", *European Journal of Marketing*, Vol. 18, No. 4, pp. 36-44, 1984.

[3] J. Hadden, A. Tiwari, R. Roy and D. Ruta, "Computer Assisted Customer Churn Management: State-of-the-Art and Future Trends", *Computers and Operations Research*, Vol. 34, No. 10, pp. 2902-2917, 2007.

[4] Emilia Huong Xuan Nguyen, "Customer Churn Prediction for the Icelandic Mobile Telephony Market", Master Thesis, Department of Industrial Engineering, University of Iceland, pp. 1-109, 2011.

[5] Robert E. Schapire, "The Boosting Approach to Machine Learning an Overview", *Proceedings of International Conference on Nonlinear Estimation and Classification*, pp. 1-12, 2003.

[6] S. Karuppaiah and N.P. Gopalan, "Enhanced Churn Prediction using Stacked Heuristic Incorporated Ensemble Model", *Journal of Information Technology Research*, Vol. 14, No. 2, pp.174-186,2021.

[7] Ammar A.Q. Ahmed and D. Maheswari, "Churn Prediction on a Huge Telecom Data using Hybrid Firefly-based Classification", *Egyptian Informatics*, Vol. 18, pp.215-220, 2017.

[8] A. Saran Kumar and D. Chandrakala, "A Survey on Customer Churn Prediction using Machine Learning Techniques", *International Journal of Computer Applications*, Vol. 154, No. 10, pp. 975-987, 2016.

[9] A. Somasundaram and U.S. Reddy, "Data Imbalance: Effects and Solutions for Classification of Large and Highly Imbalanced Data", *Proceedings of International Conference on Research in Engineering, Computers and Technology*, pp. 1-16, 2016.

[10] H. Jain, A. Khunteta and S. Srivastava, "Churn Prediction in Telecommunication using Logistic Regression and Logit Boost", *Procedia Computer Science*, Vol. 167, pp. 101-112, 2020.

[11] D. Slof, F. Frasincar and V. Matsiiako, "A Competing Risks Model-based on Latent Dirichlet Allocation for Predicting Churn Reasons", *Decision Support Systems*, Vol. 146, pp. 113541-113549, 2021.

[12] B. Lariviere and D. Van Den Poel, "Investigating the Role of Product Features in Preventing Customer Churn, by using Survival Analysis and Choice Modeling: The Case of Financial Services", *Expert Systems with Applications*, Vol. 27, pp. 277-285, 2005.

[13] Z. Jamal and R. Bucklin, "Improving the Diagnosis and Prediction of Customer Churn: A Heterogeneous Hazard Modeling Approach", *Journal of Interactive Marketing*, Vol. 20, No. 3-4, pp. 16-29, 2006.

[14] L. Calzada Infante, M. Oskarsdottir and B. Baesens, "Evaluation of Customer Behavior with Temporal Centrality Metrics for Churn Prediction of Prepaid Contracts", *Expert Systems with Applications*, Vol. 160, pp. 113553-113563, 2020.

[15] Omar Adwan, Osama Harfoushi, Hossam Faris and Nazeeh Ghatasheh, "Predicting Customer Churn in Telecom Industry using Multilayer Perceptron Neural Networks: Modeling and Analysis", *Life Science*, Vol. 11, No. 3, pp. 75-81, 2014.

[16] T. Cenggoro, R. Wirastari, E. Rudianto, M. Mohadi, D. Ratj and B. Pardamean, "Deep Learning as a Vector Embedding Model for Customer Churn", *Procedia Computer Science*, Vol. 179, pp. 624-631, 2021.

[17] N. Alboukaey, A. Joukhadar and N. Ghneim, "Dynamic Behavior-based Churn Prediction in Mobile Telecom", *Expert Systems with Applications*, Vol. 162, pp. 113779-113789, 2020.

[18] J. Zaratiegui, A. Montoro and F. Castanedo, "Performing Highly Accurate Predictions Through Convolutional Networks for Actual Telecommunication Challenges", *Proceedings of International Conference on Computer Vision and Pattern Recognition*, pp. 1-8, 2015.

[19] A. Wangperawong and C. Brun, "Churn Analysis using Deep Convolutional Neural Networks and Autoencoders", *Proceedings of International Conference on Computer Vision and Pattern Recognition*, pp. 1-8, 2016.

[20] M. Oskarsdottir, T. Van Calster and B. Baesens, "Time Series for Early Churn Detection: using Similarity-based

Classification for Dynamic Networks", *Expert Systems with Applications*, Vol. 106, pp.55-65, 2018.

[21] Wai Ho Au, Keith C.C. Chan and Xin Yao, "A Novel Evolutionary Data Mining Algorithm with Applications to Churn Prediction", *IEEE Transactions on Evolutionary Computation*, Vol. 7, No. 6, pp. 1-15, 2003.

[22] S. Kostic, M. Simic and M. Kostic, "Social Network Analysis and Churn Prediction in Telecommunications using Graph Theory", *Entropy*, Vol. 22, No. 7, pp. 753-767, 2020.

[23] S. Hoppner, E. Stripling, B. Baesens, S. Broucke and T. Verdonck, "Profit Driven Decision Trees for Churn Prediction", *European Journal of Operational Research*, Vol. 284, No. 3, pp. 920-933, 2020.

[24] S. Maldonado, J. Lopez and C. Vairetti, "Profit-based Churn Prediction-based on Minimax Probability Machines", *European Journal of Operational Research*, Vol. 284, No. 1, pp. 273-284, 2020.

[25] E. Stripling, S. Vanden Broucke and K. Antonio, "Profit Maximizing Logistic Model for Customer Churn Prediction using Genetic Algorithms", *Swarm and Evolutionary Computation*, Vol. 40, pp. 116-130, 2018.

[26] G. Lanckriet, L. Ghaoui and C. Bhattacharyya, "A Robust Minimax Approach to Classification", *Journal of Machine Learning Research*, Vol. 3, pp. 555-582, 2003.

[27] Churn in Telecom's Dataset, Available at https://www.kaggle.com/becksddf/churn-in-telecoms-dataset, Accessed at 2020.