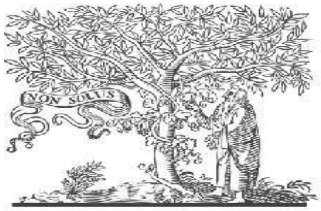
International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

COPY RIGHT



ELSEVIER SSRN

IJIEMR Transactions, online available on 05th Apr 2023. Link :http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=Issue 04

10.48047/IJIEMR/V12/ISSUE 04/06

Title **TELECOM CHURN PREDICTION**

Volume 12, ISSUE 04, Pages: 37-44

Paper Authors
Mrs. D. Vamsi, N. Leela Tejaswini, M. Aruna, Pavan Kalyan .K, N. Gandhi Rajeev

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Telecom Churn Prediction

Mrs. D. Vamsi[1], Assistant Professor, Department of CSE,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.

N. Leela Tejaswini[2], **M. Aruna[3]**, **Pavan Kalyan .K[4]**, **N. Gandhi Rajeev[5]**
[2,3,4,5] UG Students, Department of CSE,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.
[1]d.vamsi1@gmail.com, [2]leelatejaswini007@gmail.com, [3]mogiliaruna2002@gmail.com,
[4]pavankalyan04082001@gmail.com , [5]rajeevnelapati1@gmail.com

Abstract

Customers in the telecom sector have access to a variety of service providers and can actively switch from one operator to another. In this fiercely competitive market, the telecom business has an average annual churn rate of 15 to 25 per cent. Retaining highly profitable consumers is the top business objective for many established operators. Telecom businesses must identify the consumers who are most likely to leave in order to reduce customer turnover. Using the data (features) from the first three months, the business goal is to estimate the churn in the most recent (i.e., ninth) month. Considering the normal consumer behaviour during churn will help with this endeavour. The novel approach frequently employ a Logistic Regression model and Random Forest Model are used to actively capture the business objective of predicting customer churn and also this study used PCA model to find out the most effective features by feature reduction. The accuracies acquired are 83%, 75% respectively. The reasons that the service provider can use to better and completely reduce customer turnover are also provided.

Keywords: Machine Learning, PCA, Recall, Recursive Feature Elimination, Telecom Churn.

Introduction

The telecom sector has two primarily payment methods: post-paid (where consumers pay a monthly or annual fee after using the services) and prepaid (where users pay/recharge with a set amount in advance and then use the services).In the post-paid model, customers typically contact the current operator to cancel the services when they want to switch to another operator, hence this study can immediately identify this as a case of churn. However, with the prepaid model, users can abruptly stop using the services and switch to another network, making it difficult to determine if a user has genuinely left or is only utilising the services less frequently (e.g. someone may be on a trip abroad for a month or two and then intend to resume using the services again).Therefore, predicting churn for prepaid clients is typically more important (and complex), and the term "churn" needs to be defined precisely. Additionally, prepaid plans are more popular in India and Southeast Asia

than post-paid plans are in North America and Europe. The Indian and Southeast Asian markets are the foundation of this initiative.

Churn can be characterised in a number of ways, including:

A. Revenue Based Churn:

Customers who have not used any revenue-generating services during a certain length of time, such as mobile internet, outbound calls, SMS, etc. The use of aggregate metrics is another option. For example, "clients who have generated less than INR 4 lakhs per month in total/average/median revenue."

The biggest flaw in this definition is that some consumers just use the services to receive calls or SMSs from their wage-earning counterparts; in plenty of other words, they don't produce any income but nevertheless use the services. For instance, many people in rural locations only get calls from their siblings who work in cities.

B. Usage Based Churn:

Customers that have not made any calls, used the internet, or made any other outgoing or incoming usage over a period of time (usage-based churn).

Once they have stopped utilizing the services for a period of time, this definition may have the drawback that it may be too late to take corrective measures to retain the client.

For instance, if we use a "two-month zero usage" definition of churn, projecting churn may be meaningless because the client will have already gone to a different operator by that point.

In this project, churn will be defined using the usage-based definition.

High Value Churn:

About 80% of revenue in the Indian and Southeast Asian markets comes from the top 20% of clients (called high-value customers). Therefore, if high-value client churn can be lowered, large revenue leakage can be lowered.

In this project, this study will exclusively forecast churn on high-value customers and identify high-value customers based on a certain criterion (described below).

The dataset comprises customer-level data for the months of June, July, August, and September along a four-month period. The corresponding month codes are 6, 7, 8, and 9.Using the data (features) from the first three months, the business goal is to estimate the churn in the most recent (i.e., ninth) month. Understanding the normal consumer behaviour during churn will help with this endeavour.

Literature Survey:

As a part of literature survey, we have considered mainly to review 10 papers which are already done on this churn prediction. The papers are as follows:

In paper [3], the authors combined deep learning models with AI to predict the churn. They acquired an accuracy of 86% and F1 score of 78%.

In paper [1], authors investigated the applicability of competing risk model and LDA to predict the churn.

Paper [2], presents a Meta heuristic based churn prediction technique that performs churn prediction on huge telecom data. The firefly algorithm depicted the accuracy of 86.38%.

Dynamic behaviour of churn is calculated in the paper [4] using the models like CNN and LSTM for automatic feature selection. Manual feature selection is done by using models like Statistics model.

The authors of paper [5] presented a comparative study on the most popular machine learning methods applied to the challenging problem of customer churning prediction in the telecommunications industry.

Ensemble models like Random Forest and XGBoost can be used for the feature selection strategies as used in papers [7] and [10]. Whereas the papers [6] and [8] discussed about the empirical decisions on opinion leaders who greatly influenced the other type of users of mobile firm.

Paper [9] considered a similar approach to machine learning by using the combination of Logistic regression and Logit boost algorithms to find out the churn rate.

Problem Identification:

After observing all the different approaches followed for the identification of churn, the main conclusion drawn is that almost all of them are mainly focused on increasing the accuracy. Whereas the main business objective lies in sensitivity/recall which is mainly the no. of positive outcomes after training the model. This study focused on the recall of the outcomes taken in the dataset. The dataset considered is a publicly available dataset which is similar to the large telecom firm in Indian and South Asian countries. So, the main goal of this article is to increase the recall/sensitivity of the dataset given. Inorder to do that, considered models are PCA (Principal Component Analysis) for the feature reduction and dimensionality reduction. Machine learning models like Logistic Regression and Random Forest Classifier are used. Also both the models are implemented without using PCA as well. In that process, RFE (Recursive Feature Elimination) is used in both of the models.

Methodology:

So in order to increase the sensitivity i.e. the no. of positive outcomes, we used the dataset with 99999 records and 226 features. The dataset contains the data in a four months window format. The four months are divided into three phases: Good, Action and Churn phases. First three months are considered as the Good phase where the customers behave well. Next month is divided into the action

phase in which the customer tends to behave abnormally. In action phase, the customer had already left the firm. As all the features are not required for the analysis, to perform feature reduction PCA (Principal Component Analysis) is used.

Data Preparation:

For this issue, it's critical to perform the data preparation stages listed below:

Obtain fresh characteristics: The ability to distinguish between excellent and bad models using good features makes this one of the most crucial steps in the data preparation process.

Filter High-value customers: We simply need to forecast turnover for high-value clients. Those who have recharged for an amount greater than or equal to X, where X is the 70th percentile of the typical recharge amount for the first two months.

Mark churners and strip away churn phase characteristics: Now, based on the fourth month, tag the churned consumers as follows: Those who are in the churn phase but have not made any calls (incoming or outgoing) OR even accessed mobile internet once.

After the data preparation, EDA is performed in which five new columns are created for the analysis where churn is unbothered as they are deleted before actual implementation. The five new columns created are:

decrease_mou_action,
decrease_rech_num_action,
decrease_rech_amt_action,
decrease_arpu_action,
decrease_vbc_action.

For EDA, both univariate and bivariate analysis are considered in Fig 1 and Fig 2

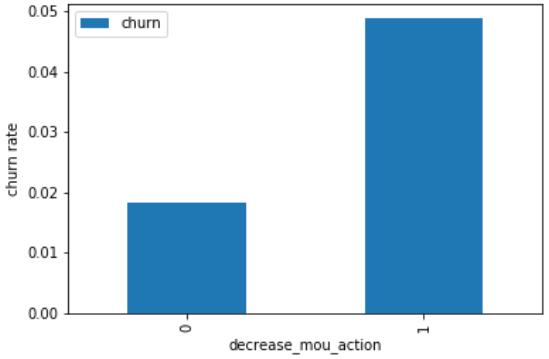


Fig 1: Univariate Analysis of EDA.

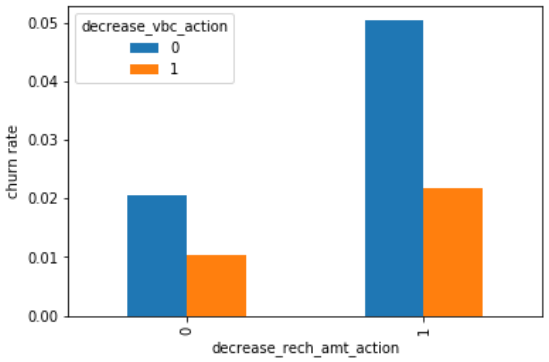


Fig 2: Bivariate Analysis of EDA

The models used for acquring an increase in the recall/senstitivity is as follows:
Principal Component Analysis (PCA) for dimensionality and feature reduction. PCA is a unsupervised learning technique used for the dimentionality reduction. PCA takes the original features and forms the linear combination of features which are known as components. Originally these components are given by the user itself. But in this study to find out the no. of components to be considered, cummulative variance of the components.

In the outputs, top 60 components are considered as they cover the 90% variance of the dataset elements.

Fig 3 explains so. After PCA, machine learning models as logistic regression and random forest classifier are used. This study considered four main models. Logistic Regression with PCA, Logistic Regression without PCA, Random forest claasifier with PCA, Random forest classifier without PCA.

Logistic Regression is mainly used when there is a binary classification of elements such as yes or no events. As this study mainly consideres churn or not churn of customers, logistic regression is used.

Random Forest classifier is mainly used to increase the accuracy by decision tree classification which mainly focuses on the accuracy increment. While trying to stay close to the business objective, Random forest classifier is used for accuracy.

Implementation:

After the EDA, actual implementation of PCA is done. Before that the class imbalance is treated using the SMOTE (Synthetic Minority Oversampling Technique) where the minority sample is taken and the equal replicas are created as samples which clear the class imbalance but cause the oversampling. Smote is taken into consideration because it mainly operates on features which is further needed here. After the Smote, feature scaling is performed before applying the PCA because the model is outperformed on the scaled version. The study will try to use ways to tackle class imbalance because the rate of churn is normally modest (between 5 and 10%; this is known as class-imbalance).To create the model, the following steps might be considered:

- Preprocess data (convert columns to appropriate formats.
- Conduct the necessary exploratory investigation to uncover beneficial insights.
- Obtain fresh characteristics.
- Utilize PCA to lower the number of variables

After using PCA, the components are reduced to about 60 from the existing 226. PCA converts the original features into principal components which are basically the linear combination of the original features.

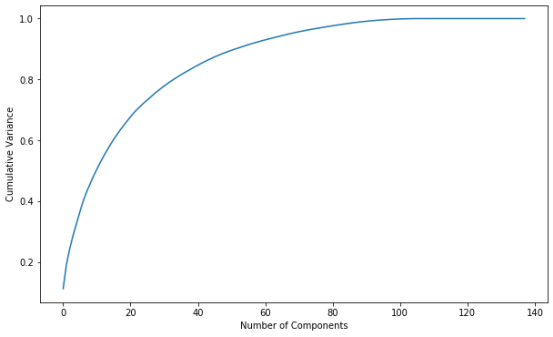


Fig 3: Graph for cumulative variance and no. of components.

Now only the components that are considered after applying pca are taken as train set. Test set is transformed using the feature scaling. Train and Test sets are divided in 80:20 ratio. Further, Models as Logistic Regression and Random Forest Classifier are applied.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Both the models are given as with pca and without pca.

Logistic Regression: To perform logistic regression with PCA, the steps followed are as follows:

1. Data Standardization: It is important to standardize the data before performing PCA.
2. Perform PCA: Use PCA to reduce the dimensionality of the data.
3. Split the data: Split the data into training and testing sets.
4. Fit the logistic regression model: Fit a logistic regression model using the top principal components as predictor variables.
5. Evaluate the model: Evaluate the performance of the logistic regression model.

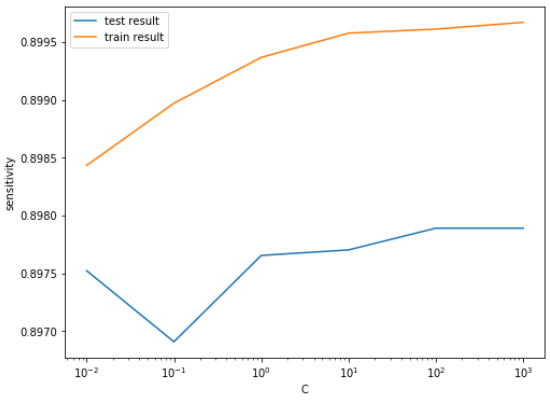


Fig 4: Graph between train and test results for logistic regression with pca.

To perform the logistic regression without using PCA, this study used Recursive Feature Elimination (RFE). RFE, unlike PCA, eliminates the least important features for every recursion. The n number of features to be remained is also specified in the model syntax along with the classification model to be used.

Random Forest Classifier:

To perform random forest classifier with PCA, the steps followed are as follows:

1. Standardize the data by subtracting the mean and scaling to unit variance.
2. PCA is applied to the standardized data.
3. Determine the number of principal components to keep based on the percentage of variance explained or by cross-validation.
4. Transform the standardized data to the reduced feature space.
5. Train a Random Forest model on the transformed data.
6. Evaluate the performance of the Random Forest model on a test set.

To perform random forest classifier, without using PCA, again we use recursive feature elimination. Here, as the random forest classifier takes more than a valid amount of time, a step factor for rfe is used. Step factor skips the given n no. of steps while performing recursions. Here the considered step factor is 6 or 7. As the no. of factors here after the outlier treatment are of almost 136 features, 6 or 7 is considered as the skip or step factor.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

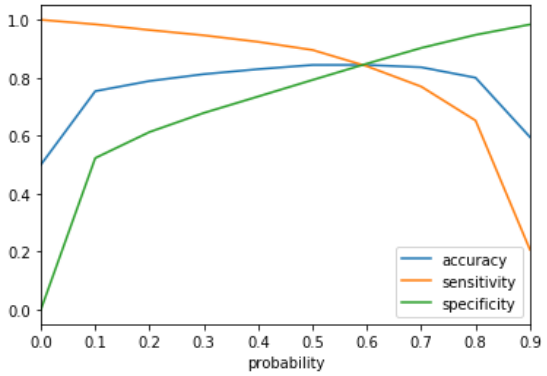


Fig 5: Metrics of Random Forest Classifier using RFE.

Results and Conclusion:

After applying the four different models on the concerned data set, the results are depicted in the following Table 1:

Model/Metric	M1	M2	M3	M4
Accuracy	83%	79%	84%	78%
Recall	89%	75%	89%	82%
Specificity	83%	80%	79%	78%

Table 1: Metrics Measurement for all the used models.

In the above Table 1, models Logistic Regression with PCA, Random forest classifier with PCA, Logistic Regression without PCA, and Random Forest classifier without PCA are labelled as M1, M2, M3, and M4 respectively.

By observing all the models, this study concludes that, Logistic Regression model performs equally with and without using PCA. It acquired the same recall in both the cases. Whereas the accuracy is decreased in PCA model. Hence, either of the models can be considered. Graphs of metrics occurred are specified in Fig 4 and Fig 5.

While using Random Forest classifier, the model trained with PCA observed accuracy increment. Whereas, the model trained with unrefined dataset observed a rise in recall value.

Limitations and Future Scope:

The novel approach used PCA for feature reduction which almost reduced the components to half an amount. But the other least important half can also contain important elements. This serves as a limitation as the accuracy might be incremented if that feature is used. Many other techniques like Forward elimination, LDA etc., can be used for feature selection and deep learning algorithms can be used to find out the churn predictions and the increment of accuracy and recall as well.

References:

[1] Slof, D., Frasincar, F., et.al. (2021). A competing risks model based on LDA for predicting churn reasons. *Decision Support Systems*, *146*, 113541.

[2]Ahmed, A. A., et.al (2017). Churn prediction on huge telecom data using the famous hybrid firefly based classification. *Egyptian Informatics Journal*, *18*(3), 215-220.

[3]Cenggoro, T. W., et.al. (2021). Deep learning as a vector embedding model for customer churn. *Procedia Computer Science*, *179*, 624-631.

[4]Alboukaey, N., et.al. (2020). Dynamic behavior based churn prediction in mobile

telecom. *Expert Systems with Applications*, *162*, 113779.

[5]Vafeiadis, T., et.al.(2015). Comparisons of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory*, *55*, 1-9.

[6] Chen, C. P., et.al. (2018). Employing a data mining approach for identification of mobile opinion leaders and their content usage patterns in large telecommunications of datasets. *Technological Forecasting and Social Change*, *130*, 88-98.

[7] Idris, A., Rizwan, M., & Khan, A. (2012). Churn prediction in telecom using Random Forest and PSO based data balancing in combination with various feature selection strategies. *Computers & Electrical Engineering*, *38*(6), 1808-1819.

[8] Uner, M. M., Guven, F., & Cavusgil, S. T. (2020). Churn and loyalty behavior of Turkish digital natives: Empirical insights and the managerial implications. *Telecommunications Policy*, *44*(4), 101901.

[9] Jain, H., Khunteta, A., & Srivastava, S. (2020). Churn prediction in telecommunication using logistic regression and logit boost. *Procedia Computer Science*, *167*, 101-112.

[10] Shrestha, S. M., & Shakya, A. (2022). A Customer Churn Prediction Model using XGBoost in Nepal. *Procedia Computer Science*, *215*, 652-661.

[11] Hung, S. Y., Yen, D. C., & Wang, H. Y. (2006). Applying data mining to telecom churn management. *Expert Systems with Applications*, *31*(3), 515-524.

[12] Karahoca, A., & Karahoca, D. (2011). GSM churn management by using fuzzy c-means clustering and adaptive neuro fuzzy inference system. *Expert Systems with Applications*, *38*(3), 1814-1822.

[13] Sulikowski, P., & Zdziebko, T. (2021). Churn factors identification from real-world data in the telecommunications industry: case study. *Procedia Computer Science*, *192*, 4800-4809.

[14] Verbeke, W., Dejaeger, et.al. (2012). New insights into churn prediction in the telecommunication sector: A profit driven data mining approach. *European journal of operational research*, *218*(1), 211-229.

[15] Al-Debei, M. M., Dwivedi, Y. K., & Hujran, O. (2022). Why would telecom customers continue to use mobile value-added services?. *Journal of Innovation & Knowledge*, *7*(4), 100242.

[16] Chen, S. H. (2016). The gamma CUSUM chart method for online customer churn prediction. *Electronic Commerce Research and Applications*, *17*, 99-111.

[17] Maldonado, S., López, J., & Vairetti, C. (2020). Profit-based churn prediction based on minimax probability machines. *European Journal of Operational Research*, *284*(1), 273-284.