

# An extended PPFCM- ANN model for telecommunication customer retention

J. Vijaya, CSE dept, VIT- AP University, Amaravati, Andrapradesh, India, vijaya.stephen@vitap.ac.in  
Hussian Syed, CSE dept, VIT- AP University, Amaravati, Andrapradesh, India, hussain.syed@vitap.ac.in

## Article Info

Volume 82

Page Number: 8796 - 8801

Publication Issue:

January-February 2020

## Abstract:

The term 'Churn' indicates a situation when an existing customer shows no longer interest to continue his or her participation in utilizing the services of a company. Hence Churn Prediction plays a crucial dynamic role that paves the way for the sustainable growth of the organization importantly in the ever-challenging telecommunication industry. This article intends to propose a framework to forecast a customer by hybrid probabilistic possibilistic fuzzy C-means clustering (PPFCM) along with Ensemble classification (PPFCM-Ensemble) techniques. This paper involves two modules: (1) PPFCM clustering techniques used for clustering module (2) Ensemble classifiers used for customer forecasting component module. During the training process, the train dataset is assembled into groups, with the assistance of PPFCM clustering techniques. The acquired segmented data is utilized in the following ensemble classifier and this mixture model is utilized for forecasting churn prediction process. During the testing procedure, the segmented test information selects the most precise ensemble classifier which relates to the nearest group of the test information, as indicated by least distance. At last, to forecast the churn customer using the proposed hybrid PPFCM-Ensemble model. Three different experiments are performed, it is proved that the proposed Hybrid PPFCM-Ensemble Model affords maximum accuracy in comparison with any solitary models.

## Article History

Article Received: 5 April 2019

Revised: 18 Jun 2019

Accepted: 24 October 2019

Publication: 08 February 2020

**Keywords:** Customer churn, Telecommunication, Clustering, Classification, PPFCM clustering, Ensemble classification, Hybrid model, Bagging, Boosting, Random subspace.

## I. INTRODUCTION

For the development of the clientele, the growth rate indicating the number of new customers' needs to be improved than the churn rate. Hence, using enhanced data mining techniques a viable churn prediction system is designed [1]. The ensuing goal of every organization is to preserve the prevailing customer base, because accumulating new customers may involve enormous money investment, time consuming and may pile up more of human resource [3]. The technology era in which telecommunication extends as a key industry for the positive economic impact. But customer churn is very high in telecommunication sector like nearly 27% of valuable customer's defection happens in every year [4], it is very important to predict the customer churn

using modern data mining and prediction technologies. Such technologies can be implemented using fuzzy logic, neural networks, support vector machines, regression analysis, decision tree, naïve Bayes and other rule based predicting mechanisms [5,8,11]. But it was produced unsatisfied result, hence it is interest of many industries and research people to focus on active churn prediction research. This article intends to propose a framework to forecast a customer by hybrid probabilistic possibilistic fuzzy C-means clustering (PPFCM) along with Ensemble classification (PPFCM-Ensemble) techniques. This paper involves two modules: (1) PPFCM clustering techniques used for clustering module (2) Ensemble classifiers used for customer forecasting component module. During the training process, the train dataset is assembled into

groups, with the assistance of PPFCM clustering techniques. The acquired segmented data is utilized in the following ensemble classifier and this mixture model is utilized for forecasting churn prediction process. During the testing procedure, the segmented test information selects the most precise ensemble classifier which relates to the nearest group of the test information, as indicated by least distance. At last, to forecast the churn customer using the proposed hybrid PPFCM-Ensemble model.

## II. LITERATURE SURVEY

Ever blooming data mining techniques consists of patter analysis, pattern prediction, it finds the most significant trends, patterns, facts, and hidden anomalies in the system. If no soft computing and data mining techniques the facts may left unnoticed. So, data mining techniques play a vital role in success of every firm [3, 6, 8, 10]. Various data mining techniques like Neural Networks, Legit Regression, DT and Boosting techniques were implemented utilizing the churn prediction data set consisting of 47000 domestic users from U.S. The attributes like credit, billing and complaint history along with customer's application are included in the data set [3]. A feature selection-based technique applied on Real Ada-boost was stated in [6,7]. In this work churn prediction and churn management were illustrated undertaking various attributes or features selection techniques. Even research was made on landline telecommunication, but landline user data was not available fully and in the appropriate format. Huang et al. (2012) enumerated new attribute set for customers using landline. Various attributes like Information of grants, Service orders, Demographic profiles, the historical information about payment of bills, Henley segments, all call details were the features used to design CCP models. The classification techniques like Evolutionary Data Mining Algorithm (EDMA), Decision Tree (DT), Logistic Regression (LR), Linear Classification, Naïve Bayes (NB), Multi-Layer Perceptron (MLP), Support Vector Machines

(SVM) and Multilayer Perceptron Neural Network (MPNN) were involved for designing the prediction models [3]. It is found that extracted new features gave increased accuracy than that of existing features. The proposed system performance is measured using true and false positive rates. While analyzing the performance of the system using above mentioned classification methods, six classification methods performed very well except evolutionary mining technique [3]

## III. PROPOSED MODEL

Fig1 illustrates our well proposed system for the customer churn prediction. The Duke University data set is collected and the needed data cleansing techniques along with data pre-processing is carried out [2]. To propose a framework to forecast a customer by hybrid probabilistic possibilistic fuzzy C-means clustering (PPFCM) along with Ensemble classification (PPFCM-Ensemble) techniques. This paper involves two modules: (1) PPFCM clustering techniques used for clustering module (2) Ensemble classifiers used for customer forecasting component module. During the training process, the train dataset is assembled into groups, with the assistance of PPFCM clustering techniques. The acquired segmented data is utilized in the following ensemble classifier and this mixture model is utilized for forecasting churn prediction process. During the testing procedure, the segmented test information selects the most precise ensemble classifier which relates to the nearest group of the test information, as indicated by least distance. At last, to forecast the churn customer using the proposed hybrid PPFCM-Ensemble model.

### A. Data Collection and Cleaning

Tera Data Center focusing on CRM at Duke University consists of three different types of data [2]. Primarily or the calibration data set is considered as this experimentation. The data set totally consists of nearly 1, 00,000 samples and 172 attributes in combination with the class labels in the data set. In

the considered data set out of 172 attributes, there are about 137 numerical attributes and 35 nominal attributes. This considered data set has 20% of

missing values. It is also well labeled that nearly 50438 consumers are found to be non-churners and churning customers count is found to be 49562 [9].

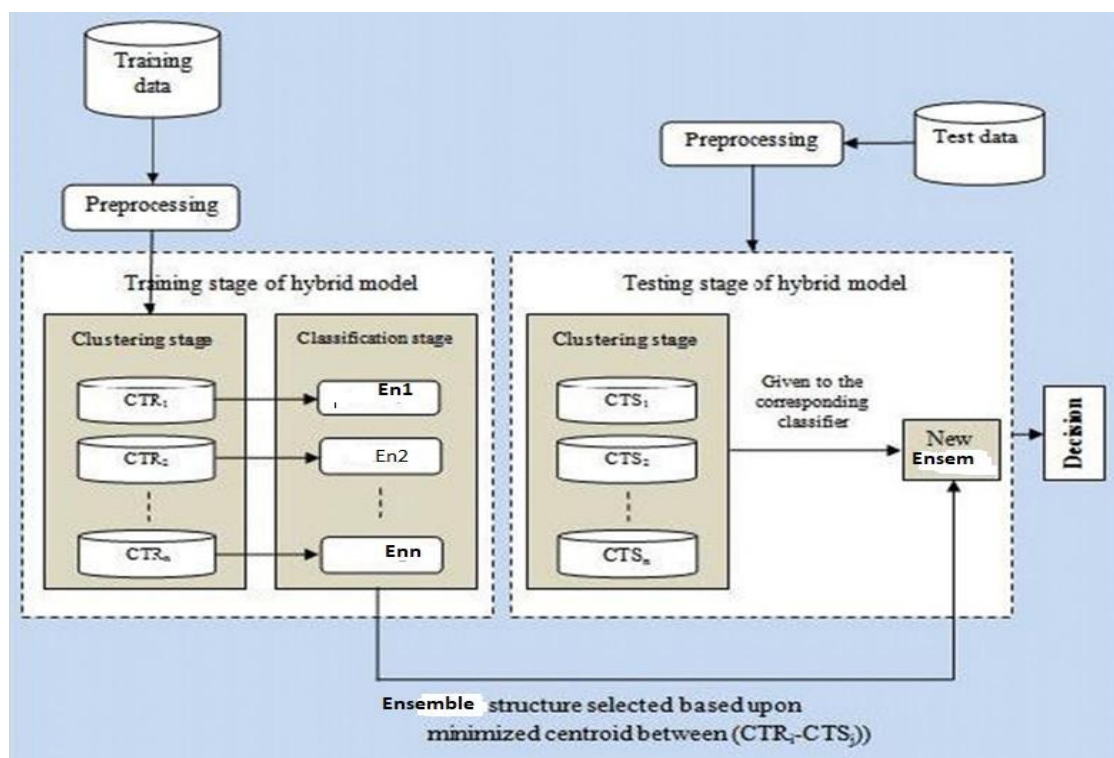


Fig 1: Enhance churn prediction method

### B. Data Cleaning and Splitting

In the initial preprocessing phase attributes like Customer ID and last\_swap features are removed so that attributes size is reduced to 170 (136 numeric and 34 nominal). The missing values of numerical attributes are replaced by mean value and the missing values of nominal attributes are replaced with the mode of the attributes. Also, the clustering algorithms use only numerical values, all the strings in the attributes set are converted into numeric values. In the further step normalization is performed using min-max normalization. The preprocessed data is split into training and test part based on number of churn percentage (Refer paper 9 section 4.1)

### C. PPFCM Clustering algorithm

In this paper, to propose and design a system which can involve in probable prediction of customer

Fuzzy C-Means Clustering Methodology (PPFCM) (Refer [9]) in combination with enhanced classification. This system effectively improves the prediction accuracy when compared with existing PPFCM-ANN model [9]. PPFCM algorithm is extended version of PFCM algorithm. In PPFCM clustering algorithm to generate a new membership function which is the multiplication of proposed newly generated probability values with membership function of existing PFCM clustering algorithm [9].

## IV. EXPERIMENTS AND RESULT ANALYSIS

Three different experiments are performed in which the primary experiment evaluates the standard classification results are compared to existing PPFCM- ANN based classification which is illustrated in the earlier paper [9]. In second experiment evaluates the proposed PPFCM-Ensemble based classification hybrid models. In third experiment evaluates the proposed PPFCM-

Ensemble based classification hybrid models are compared to existing hybrid models which is illustrated in the earlier paper [9]. Such that the proposed Hybrid PPFCM-Ensemble Model affords maximum accuracy in comparison with any solitary models.

#### A. Performance of Single Classifiers Vs Existing PPFCM-ANN model

The primary experiment evaluates the standard classification results such as DT, KNN, SVM, NB,

and ANN is compared to existing PPFCM-ANN. The accuracy obtained from the above methods are figured in Fig2 which is illustrated in the earlier paper with cluster size 2 [9]. Here number of churn percentage is 50 that means training data has 24781 churn customer and 25219 non churn customer and test data have 50000 customers. It can be inferred from the fig that PPFCM-ANN improves the accuracy considerably.

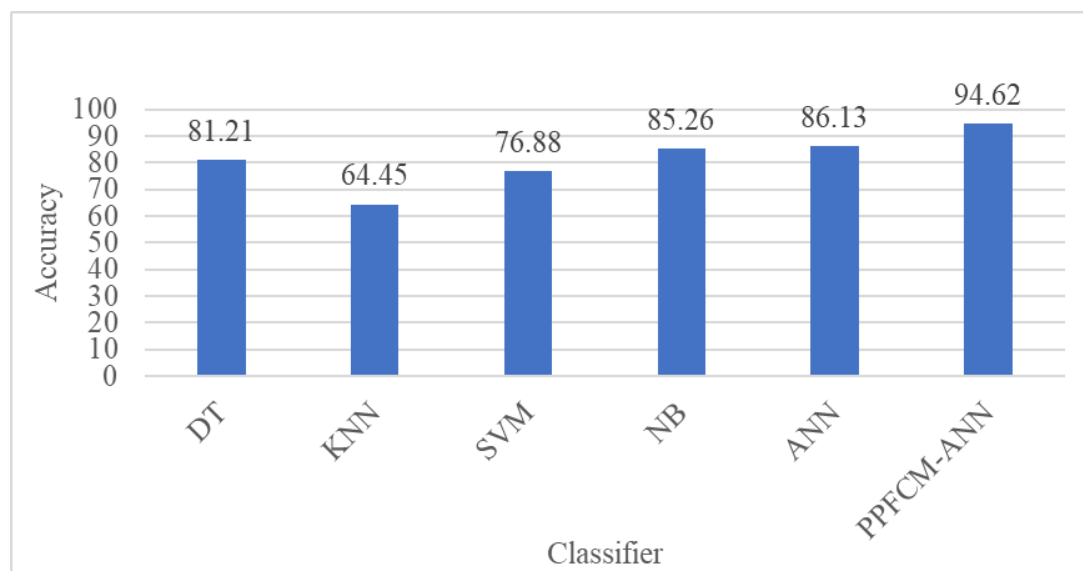


Fig 2: The accuracy variation of single classifier Vs PPFCM-ANN [9]

#### B. Performance of PPFCM clustering along ensemble classification algorithm

The second experiment evaluates the hybrid classification results such as PPFCM-bagging, PPFCM-boosting, PPFCM-Random Subspace. The accuracy obtained from the above methods are figured in Fig3 with cluster size 2. Here number of churn percentage is 50 that means training data has 24781 churn customer and 25219 non churn customer and test data have 50000 customers. It can be inferred from the fig that PPFCM-Boosting improves the accuracy considerably.

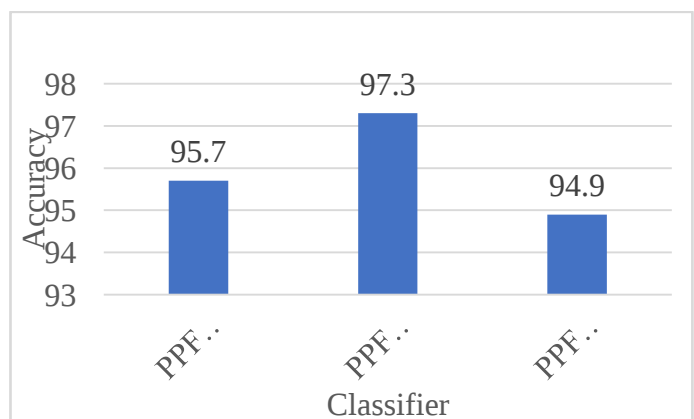
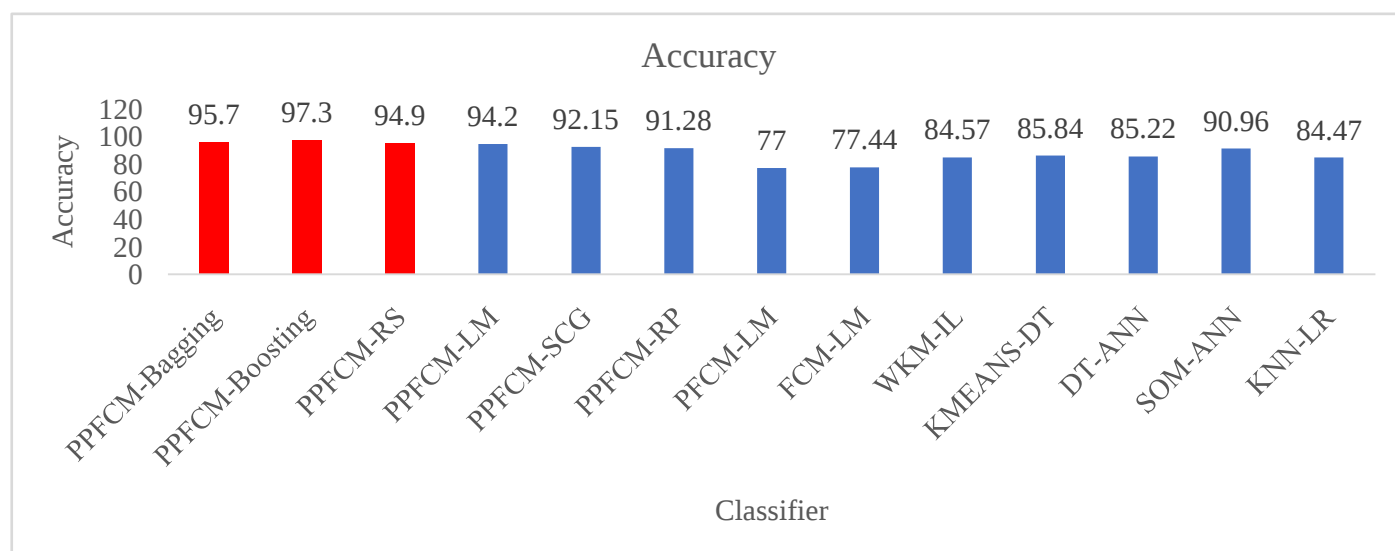


Fig 3: The accuracy variation of proposed PPFCM-Ensemble Algorithm

### C. Performance of Proposed PPFCM- enhanced classifier Vs Existing Techniques.

Final experiment substantiates the accuracy comparison between the proposed model (PPFCM-ENSEMBLE) and existing hybrid classifiers such as PPFCM-LM, PPFCM-SCG, PPFCM-RP, PFCM-LM, FCM-LM [9], WKM-IL, KMEANS-DT, DT-ANN, SOM-ANN, KNN-LR [4, 12-15]. The

accuracy obtained from the above methods are figured in Fig4 with cluster size 2. Here number of churn percentage is 50 that means training data has 24781 churn customer and 25219 non churn customer and test data have 50000 customers. It can be inferred from the fig that PPFCM-Boosting improves the accuracy considerably.



## V. CONCLUSION

In growing and ever challenging telecommunication sector churn prediction has become a vital process for the continuous growth of the organization. In this paper hybrid PPFCM with enhanced hybrid model is deployed for churn prediction in telecommunication sector. Three different experiments are performed in which the primary experiment consists of the classification result; the secondary experiment evaluates the ensemble classification result hybrid with proposed PPFCM clustering. Such that the proposed Hybrid PPFCM with enhanced classifier affords maximum accuracy in comparison with any solitary models. The third experiment evaluates the ensemble classification result hybrid with proposed PPFCM clustering is compared with existing hybrid classifiers.

## VI. REFERENCES

1. Bose I, Chen X (2009) Hybrid models using unsupervised clustering for prediction of customer churn. *J Organ Comput Electr Commer* 19(2):133–151
2. Duke University (2005) Case studies, presentations and video modules. <http://www.fuqua.duke.edu/centers/ccrm/datasetsdownload.html#data>
3. Huang, B., Kechadi, M. T., Buckley, B. (2012). Customer churn prediction in telecommunication. *Expert Systems with Applications*, 39(1), 1414-1425.
4. Huang, Y., and Kechadi, T. (2013). An effective hybrid learning system for telecommunication churn prediction. *Expert Systems with Applications*, 40 (14), 5635–5647.

5. Hudaib, A., R. Dannoun, O. Harfoushi, R. Obiedat and H. Faris (2015). Hybrid data mining models for predicting customer churn. *International Journal of Communications, Network and System Sciences*, 8(05), 91.
6. Idris, A., Khan, A., Lee, Y. S. (2013). Intelligent churn prediction in telecom: employing mRMR feature selection and RotBoost based ensemble classification. *Applied intelligence*, 39(3), 659-672.
7. Inbarani, H.H., M. Bagyamathi and A.T. Azar (2015). A novel hybrid feature selection method based on roughest and improved harmony search. *Neural Computing and Applications*, 26(8), 1859–1880.
8. Keramati A, Jafari-Marandi R, Aliannejadi M, Ahmadian I, Mozaffari M, Abbasi U (2014) Improved churn prediction in telecommunication industry using data mining techniques. *Appl Soft Comput* 24:994–1012
9. Sivasankar, E., and Vijaya, J. (2018). Hybrid PPFCM-ANN model: an efficient system for customer churn prediction through probabilistic possibilistic fuzzy clustering and artificial neural network, *Neural Computing & Application*. <https://doi.org/10.1007/s00521-018-3548-4>
10. Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G., Chatzisavvas, K. C. (2015). A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory*, 55, 1-9.
11. Vijaya. J and E.Sivasankar and S.Gayathri (2018). Fuzzy Clustering with Ensemble Classification techniques to improve the Customer Churn Prediction in Telecommunication Sector. *Neural Computing and Applications*, DOI: 10.1007/978-981-13-1280-9\_25.
12. Hung SY, Yen DC, Wang HY (2006) Applying data mining to telecom churn management. *Expert Syst Appl* 31(3):515–524
13. Lee J, Lee J (2006) Customer churn prediction by hybrid model. *Adv Data Min Appl* 4091:959–966
14. Tsai, C. F., Lu, Y. H. (2009). Customer churn prediction by hybrid neural networks. *Expert Systems with Applications*, 36(10), 12547-12553
15. Zhang Y, Qi J, Shu H, Cao J (2007) A hybrid KNN-LR classifier and its application in customer churn prediction. In: *IEEE International conference on systems, man and cybernetics, 2007, ISIC*. IEEE, pp 3265–3269