

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/333650567>

# An Effective Classifier for Predicting Churn in Telecommunication

Article in Journal of Advanced Research in Dynamical and Control Systems · June 2019

CITATIONS

6

READS

842

2 authors:



Beschi Raja

Sri Krishna College of Technology

16 PUBLICATIONS 64 CITATIONS

[SEE PROFILE](#)



Pamina Jeyakumar

Sri Krishna College of Technology

7 PUBLICATIONS 51 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



An Effective Classifier for Predicting Churn in Telecommunication [View project](#)



Networks [View project](#)

# An Effective Classifier for Predicting Churn in Telecommunication

*J. Pamina, Assistant Professor, Sri Krishna College of Technology, Coimbatore.*

*J. Beschi Raja, Assistant Professor, Sri Krishna College of Technology, Coimbatore.*

*S. Sathya Bama, Assistant Professor, Sri Krishna College of Technology, Coimbatore.*

*S. Soundarya, Assistant Professor, Sri Krishna College of Technology, Coimbatore.*

*M.S. Sruthi, Assistant Professor, Sri Krishna College of Technology, Coimbatore.*

*S. Kiruthika, Assistant Professor, Sri Krishna College of Technology, Coimbatore.*

*V.J. Aiswaryadevi, Assistant Professor, Sri Krishna College of Technology, Coimbatore.*

*G. Priyanka, Assistant Professor, Sri Krishna College of Technology, Coimbatore.*

**Abstract---** In recent days, telecom industry plays a major role in our daily life. The proliferation of telecommunication industry becomes very difficult for the service providers to survive in the market. To stabilize in this field, the service providers have to be aware of the features that make the customer to churn. The proposed predictive model identifies the traits that highly influence customer churn, with the help of machine learning techniques like KNN, Random Forest and XG Boost. IBM Watson dataset has been analysed to forecast the churn. At last a comparative study has been made among the machine learning algorithm to identify the better algorithm of higher accuracy. The proposed model shows that Fiber Optic customers with greater monthly charges have higher influence for churn.

**Keywords---** Churn, Telecommunication, XGBoost, KNN, Random Forest, IBM Watson.

## I. Introduction

Rapid increase of telecom providers rises to tremendous competition and leads to churn [1][2][3]. Churn is the process of customers switching from one firm to another in given time [22] [19]. The pre-eminent reason for churn is the customer dissatisfaction [1] [24]. Retaining the existing customers is more profitable than fetching the new customers [9]. So, the telecom providers have to maximize concentration to the extant customers to avert churn [7] [8]. Further, long standing satisfied customers provoke more profit because they aren't going to churn [13] [17]. The customer churn leads to huge loss economically and affect company's status [28]. An appropriate churn prediction model is essentially needed to predict the churners [6] [23]. Many companies are concentrating on customers induvial for giving best services to them. Churn is the inevitable action in all industry [1]. In recent past, various machine learning methods are used to tackle threat in customer churn in telecom [11] [12]. Lu et al.in (2014) suggested Gentle Ada boost algorithm [22] for churn prediction, because it reaches high accuracy and provided good separation of churn models and also a comparison is made with the results of logistic regression and single logistic regression model, which proves that the logistic regression model is the better one. Wenjie et al.in (2016) proposed Semantic-Driven Subtractive Clustering Method (SDSCM) which is proven to have more clustering strength than subtractive clustering method (SCM) and fuzzy c-means (FCM). Hadoop Map Reduce frame work is used to implement this algorithm. [28]. A.keramati et al.in (2014) compared the performance of DT, ANN, KNN, SVM and shown that hybrid of all four performs well, they also proposed a methodology for extracting the influential features from the dataset[18]. Pinar kisioglu et al.in (2010) applied Bayesian belief networks to explain the causal relationships between the features that contributes to customer churn and the important features were found out and suggested promotions to reduce the churn rate [20]. Koen W.De Bock et al.in(2012) Proposed GAMensplus classification algorithm for strong classification and interpretability. The system also demonstrated the relationship between the predictors and probabilities of churn [16]. Muhammad Azeem et al.in (2018) compared the accuracy of several classifiers with fuzzy model and proved that fuzzy classifiers are more accurate in predicting customer churn dataset having noise [4]. Chih-FongTsai et al.in (2009) used neural networks and hybridised data mining technique for predicting customer churn with higher accuracy, this system combines ANN with ANN and ANN with SOM for churn higher prediction accuracy [25]. Wouter Verbeke et al.in (2010) proposed AntMiner+ and ALBA for improving the learnability providing comprehensible and accurate customer churn prediction model [27]. Zhen-YuChen et al.in (2012) proposed Novel approach known as hierarchical multiple kernel support vector machine

(HMK-SVM) combines static customer and longitudinal behavioural data to improve churn prediction, a three-phase training algorithm was implemented. [5]. Long Zha et al.in (2017) proposed New K- local maximum margin feature extraction algorithm (KLMM), they followed the fact that extracting the features of the data will reduce the dimensionality of prediction [29].T.Vafeiadis et al.in (2015) made a comparative study on a public domain dataset, next deals with improving the accuracy by using boosting algorithms. They used Monte Carlo simulation at different settings of each classification method and improved SVM-POLY with Ada boost classifier which gave higher accuracy [26]. Ying Huang et al.in (2013) proposed Novel hybrid model-based learning system, which integrates k-means clustering algorithm (unsupervised) and inductive technique (FOIL)-(supervised) for building an effective predictive model [10]. The rest of the paper contains the following session

## II. Machine Learning Classifiers

### 1. KNN Classifier

K nearest neighbour is widely used classification method. KNN classifier classifies new cases based on the similarity measures of the available cases. The similarity measure corresponds to any functions such as distance. A case is classified based on the majority of neighbourhood data points. The distance between the datapoints is calculated using the Euclidean distance function.

$$D = \sqrt{\sum_{i=1}^k (x_i - y_i)^2}$$

Where D is the distance function and x and y correspond to the datapoints.

The above distance function is applied only for continuous variables, categorical variables will use Hamming distance.

$$D_H = \sum_{i=1}^K x_i - y_i \vee$$

In the above equation if we the value of x and y are equal then the distance function will become zero, otherwise the value of  $D_H$  is one. These criteria can be tabulated as follows

| X | Y | Distance |
|---|---|----------|
| M | M | 0        |
| M | F | 1        |

### 2. Random Forest Classifier

Random forest Classifier is one of the ensemble learning method and supervised learning algorithm which is used for both classification and regression [14] [15]. This Classifier generates a group of decision trees from the part of training set randomly selected. For each sample it creates a decision tree and predict the results from it. The higher number of trees gives the greater prediction accuracy. It then forms the combination and average of all the prediction votes from various decision trees to finalise the output. The random forest algorithm is split into two stages namely

1. Creation of Random forest
2. Performing prediction from the created random forest.

#### Creation of Random Forest

The pseudo code for the creation of random forest is as follows:

1. Select “k” features from total “m” features randomly.
2. using the best split point calculate the node “d” among the “k” features.
3. Split the node into child nodes by making use of the best split.
4. Steps 1 to 3 is repeated until “i” number of nodes has been reached.
5. The forest is built by repeating the steps 1 to 4 for “n” number times to create “n” number of trees.

#### Random Forest Prediction

The trained random forest algorithm uses the below pseudo code to perform prediction.

1. Take the test features and pass the features through the rules of each decision tree created randomly to predict the outcome. Then store the predicted outcome (target).
2. The votes for each predicted target are calculated.
3. The high voted predicted target is considered as the final prediction from the random forest algorithm.

### ***XGBoost classifier***

XG Boost stands for **extreme Gradient Boosting**. XG Boost is an implementation of gradient boosted decision trees designed for speed and performance. It is a supervised and ensemble learning method that combines trees to give a more generalizable Machine Learning model boost classifier generally built many numbers of trees and averaging the results for higher prediction.

The trees are generally built one after another, in which the errors can be refined and corrected in the previous tree and focus on the improvement of prediction accuracy in next tree. Due its cache good optimization, the XG boost classifier makes good results in prediction model but it takes more training time for iteration process. XG boost classifiers handles the sparse dataset in which the missing values are handled properly. XG boost algorithm gives very good training model and performance for prediction of accuracy. The XG boost algorithm addresses one of the key problems in tree learning which is to find the best split. For this purpose, the XG boost algorithm uses the exact greedy algorithm. The exact greedy algorithm is shown in Algorithm 1.

#### ***Algorithm 1: Exact Greedy Algorithm for Split Finding***

**Input:** I, instance set of current node

**Input:** d, feature dimension  $gain \leftarrow 0$

$G \leftarrow \sum_{i \in I} g_i, H \leftarrow \sum_{i \in I} h_i$

for  $k = 1$  to  $m$  do

$G_L \leftarrow 0, H_L \leftarrow 0$

for  $j$  in sorted (I, by  $x_{jk}$ ) do

$G_L \leftarrow G_L + g_i, H_L \leftarrow H_L + h_i, G_R \leftarrow G - G_L, H_R \leftarrow H - H_L$

$score, \frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} + \frac{G^2}{H + \lambda}$

$score \leftarrow \max_{j \in [d]}$

end

end

**Output:** Split with max score

### ***Evaluation Measures***

The classifiers Churn prediction performance for various appropriate attributes.

Precision is measure of correct positive cases predicted and evaluated by the following calculation.

$$precision = \frac{TP}{TP + FP}$$

Recall is the measure of identified correct positive cases and evaluated by the following calculation

$$recall = \frac{TP}{TP + FN}$$

Only precision or recall cannot elucidate the performance of the classifier, both measures must imply good performance. As a consequence of this phenomena, the average of precision and recall is calculated as f- measure used as measure for predicting performance of classifier. F<sub>1</sub> score attain its best performance at 1 and worst at 0.

$$F - measure = \frac{2 * precision * recall}{Precision + recall}$$

The above three evaluation measures are used for KNN, RF and XGBoost classifier.

## **III. Simulation Setup**

Our aim is to analyse the important classifiers of churn prediction. To that end, we implemented our simulation in two steps using the python language and for Analysis Libraries, pandas and kerashave been used. In the first step, all tested classifiers are initially applied on the dataset. Their performance is evaluated using the accuracy score and F -measure criteria.

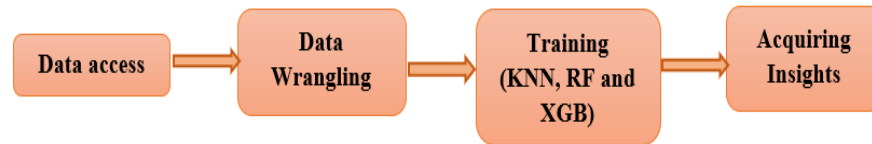


Fig. 4.1: Overall Process of KNN, RF and XGBoost Classifier

Generally, the process involved in the classifiers have four stages. Data access is nothing but giving input to the model which is the first process for analytics. Data Wrangling is a process for selecting the distribution of data which we use certain data for analysis. Training is implemented by building the models with three classifiers and test it for better accuracy. The last process is acquiring Insights in which final output are predicted and understanding the results and parameters for good prediction.

### Dataset

Telecom dataset is not publicly available more due to its customer's personal privacy. Data set for this paper obtained from IBM Watson dataset released on 2015 [30]. The data set contains 7043 instances and 21 attributes. The last attributes denote churn or not in which 5174 are not churners and 1869 are churners. The percentage of churners is 26.53% and non-churners is 73.46%. This dataset helps to figure out customer Prophecy and build retention possibilities.

### Data Pre-Processing

Data Pre-Processing is the one of the basic and important work in the sphere of machine learning. In this study, two tasks were carried out for pre-processing the data. Firstly, we have performed the missing value analysis and it is found that there no missing value present in the dataset. Next, the conversion of data types ie. Object data type to numerical data type is implemented. Total charges attribute is converted to numerical data type from object data type for the classifiers we used for analysis. We have also checked for duplication of customer id and it shows that all the rows are unique.

### Univariate Analysis

The univariate analysis is very simplest task performed for analysing the data. In this work, we used the univariate analysis for understanding the unique values of each and every features distribution to observe the objects and non-object data points.

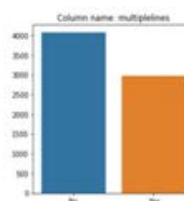
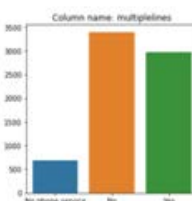
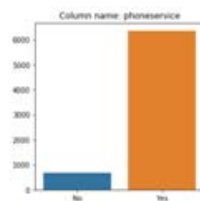


Fig. 4.2: Phone Services Fig.4.3: Multiple Lines Fig.4.4: Combination of Two

Using univariate analysis, the features are normalized based on the value of subscriber's services, there are no phone service, no and yes. Here phone service and multiple lines are primary and secondary attributes respectively. Fig. 4.2 represents the customers having phone services or not in the dataset. Fig. 4.3 indicates the number of customers having multiple lines or not and no phone service. Fig. 4.4 represents the outcome of univariate analysis which aggregates no phone service and no multiple lines.

## IV. Results and Discussion

### Data Correlation

Before building the churn, prediction model using classifiers of KNN, RF and XGBoost the correlation among the features are identified and sorted in the descending order.[32] Above table indicates the correlation values of the attributes and it is plotted as a graph (Fig 5.1) Based on this assumption, it shows clearly that there is a strong correlation between the attributes churn and Month-to-month contracts. Fibre Optic ISP and monthly charges are the second and third attributes that shows strong correlation.

Table 1: Features and Its Correlation Values

| Attributes               | Correlation values |
|--------------------------|--------------------|
| churn                    | 1.000000           |
| contract_Month-to-month  | 0.405103           |
| ISP_Fiber_optic          | 0.308020           |
| payment_Electronic_check | 0.301919           |
| Monthlycharges           | 0.193356           |
| paperlessbilling         | 0.191825           |
| Seniorcitizen            | 0.150889           |
| Streamingtv              | 0.063228           |
| streamingmovies          | 0.061382           |
| Multiplelines            | 0.040102           |
| Phoneservice             | 0.011942           |

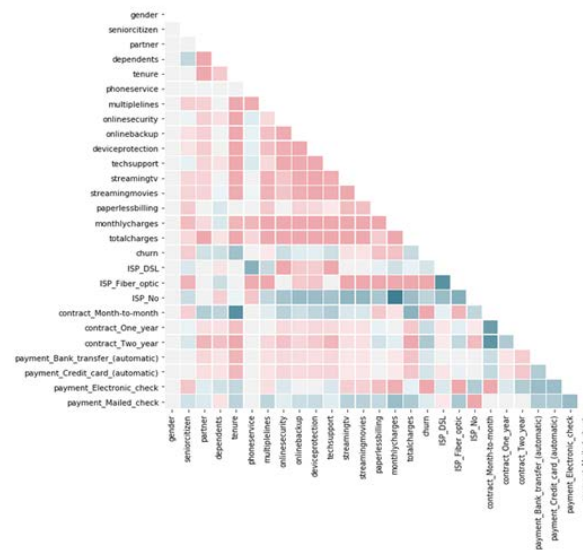


Fig. 5.1: Correlation among Features

### Comparative Analysis

The churn prediction model is created by three well known classifiers namely KNN, RF and XGBoost. The comparison of evaluation metrics such as accuracy score and F score are calculated. We have calculated the F score because accuracy alone is insufficient for higher churn prediction model. Experimental results show (Table2) that XGBoost classifier gives higher accuracy score and F score compared to KNN and RF classifiers. Thus, XGBoost classifier is used for feature selection in the proposed system.

Table 2: Results Summary of KNN, RF and XGB

| Classifier | Accuracy score | F- score     |
|------------|----------------|--------------|
| KNN        | 0.754          | 0.495        |
| RF         | 0.775          | 0.506        |
| XGB        | <b>0.798</b>   | <b>0.582</b> |

### Implementation of XGBoost

Binary Logistic regression is a baseline algorithm used here. We have used XGBoost classifier in the rest of the work for feature selection of customer churn with same data set. Memory efficiency and the training speed is

handled in XGBoost by an internal data structure called DMatrix. Based on F Score and taking 15 different features, the feature importance values are visualized as Fig.5.2.

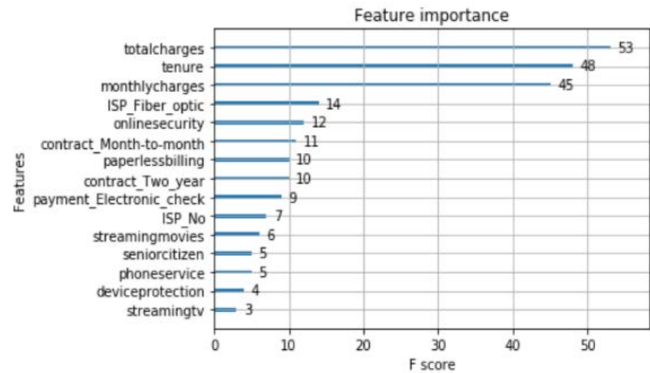


Fig.5.2 Feature Importance

Table 3: Confusion Matrix

|              | Predicted False | Predicted True |
|--------------|-----------------|----------------|
| Actual False | 1234            | 330            |
| Actual True  | 150             | 399            |

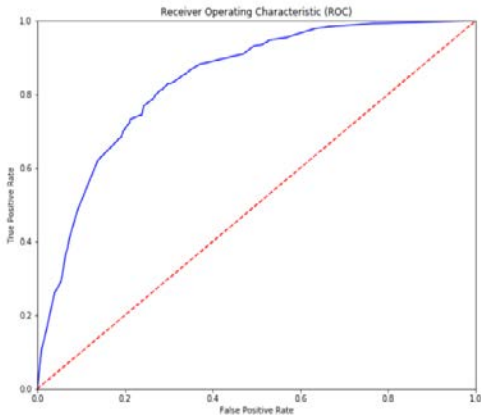


Fig.5.3: ROC



Fig.5.4: Pruned Decision Tree

Table 3 the confusion matrix calculated and it is identified that 1234 customers falls under non-churners and 399 customers falls under churners. From this, we observe that number of non-churners are more than churners. So, our classification accuracy is correctly tested. Fig 5.3 displays the ROC curve of XG Boost classifier. ROC curve is mostly used to measure the test's ability as a criterion. Generally, final model accuracy is predicted using ROC curve. [31] In the figure 5.2, the curve is stretched to true positive area which provides better prediction rates. The packages plot\_tree and graphviz are used to construct the decision tree. Theplot\_tree function takes best\_xgb (Best XBG value), num\_trees(Number of Decision trees) as parameters. The decision tree is plotted as figure.5.4.SHAPpackage is used for better visualization of XG boost classifier.

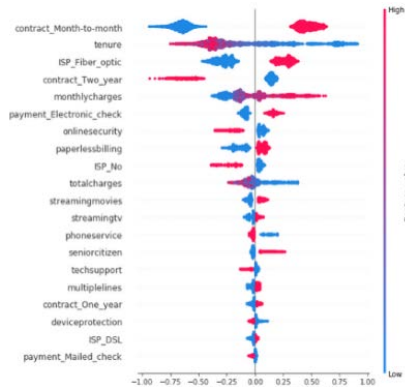


Fig.5.5: Feature Importance

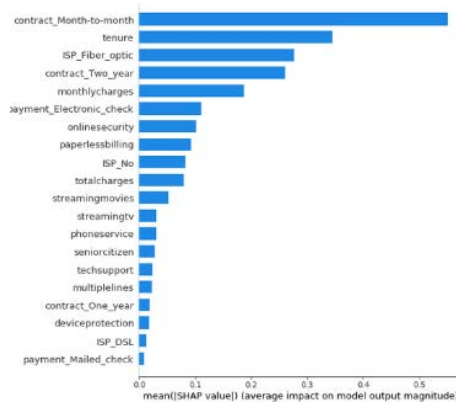


Fig.5.6: Feature Importance

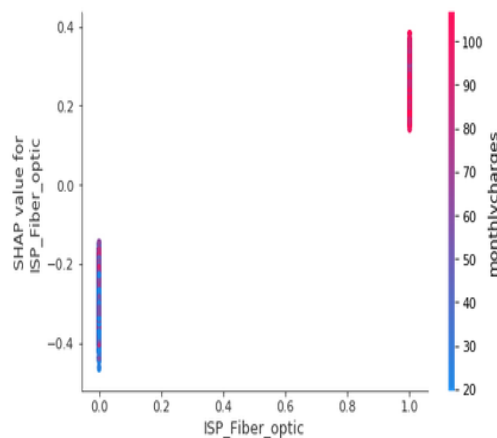


Fig 5.7: ISP Fiber Optic



We have used TreeExplainer to the XGBoost model object. After training the model, we got SHAP values. For clear explanations of entire dataset and to visualize which feature is influencing the output, the data were rotated and stacked. We also displayed the features which are influencing the output. For the Tree SHAP implantation, we have focused on the margin output model not the actual or predicted output. This indicates values units are log odd ratios. Subscriber who has large positive values are likely to churn. Rest of the attributes are used as feature importance plot in above figures. Finally, it is observed that Fiber Optic customers with greater monthly chargers are predicted as churners.

## V. Conclusion

A Comparison study has made with three prominent classifiers namely K-NN, Random Forest and XG boost for improving accuracy of customer churn prediction. The XG boost classifier performs well out of three on the publicly available telecom dataset. Next, the work was focused to identify the attribute which has higher cognition for churn using XG boost classifier. The experiment result shows that Fiber Optic customers with greater monthly charges have higher influence for churn. Anticipated directions can be to predict by hybrid of classifiers which gives high accuracy and desirable results.

## References

- [1] Ahmed, A. A. and Maheswari, D. Churn prediction on huge telecom data using hybrid firefly based classification. *Egyptian Informatics Journal* **18** (3) (2017) 215-220.
- [2] Amin, A., Al-Obeidat, F., Shah, B., Adnan, A., Loo, J. and Anwar, S. Customer churn prediction in telecommunication industry using data certainty. *Journal of Business Research* **94** (2019) 290-301.
- [3] Amin, A., Shehzad, S., Khan, C., Ali, I. and Anwar, S. Churn prediction in telecommunication industry using rough set approach. *New Trends in Computational Collective Intelligence*, 2015, 83-95.
- [4] Azeem, M., Usman, M. and Fong, A. C. M. A churn prediction model for prepaid customers in telecom using fuzzy classifiers. *Telecommunication Systems* **66** (4) (2017) 603-614.
- [5] Chen, Z. Y., Fan, Z. P. and Sun, M. A hierarchical multiple kernel support vector machine for customer churn prediction using longitudinal behavioral data. *European Journal of operational research* **223** (2) (2012) 461-472.
- [6] Chu, B. H., Tsai, M. S. and Ho, C. S. Toward a hybrid data mining model for customer retention. *Knowledge-Based Systems* **20** (8) (2007) 703-718.
- [7] De Bock, K. W. and Van Den Poel, D.. An empirical evaluation of rotation-based ensemble classifiers for customer churn prediction. *Expert Systems with Applications* **38** (10) (2011) 12293-12301.
- [8] De Caigny, A., Coussement, K. and De Bock, K. W. A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. *European Journal of Operational Research* **269** (2) (2018) 760-772.
- [9] Sivasankar, E. and Vijaya, J. Hybrid PPFCM-ANN model: an efficient system for customer churn prediction through probabilistic possibilistic fuzzy clustering and artificial neural network. *Neural Computing and Applications*, 2019, 1-20.
- [10] Huang, Y. and Kechadi, T. An effective hybrid learning system for telecommunication churn prediction. *Expert Systems with Applications* **40** (14) (2013) 5635-5647.
- [11] Huang, B., Buckley, B. and Kechadi, T. M. Multi-objective feature selection by using NSGA-II for customer churn prediction in telecommunications. *Expert Systems with Applications* **37** (5) (2010) 3638-3646.
- [12] Hung, S. Y., Yen, D. C. and Wang, H. Y. Applying data mining to telecom churn management. *Expert Systems with Applications* **31** (3) (2006) 515-524.
- [13] Idris, A., Iftikhar, A. and Ur Rehman, Z. Intelligent churn prediction for telecom using GP-AdaBoost learning and PSO undersampling. *Cluster Computing*, 2017, 1-15.
- [14] Idris, A., Rizwan, M. and Khan, A. Churn prediction in telecom using Random Forest and PSO based data balancing in combination with various feature selection strategies. *Computers & Electrical Engineering* **38** (6) (2012) 1808-1819.
- [15] Idris, A., Khan, A. and Lee, Y. S. Intelligent churn prediction in telecom: employing mRMR feature selection and RotBoost based ensemble classification. *Applied intelligence* **39** (3) (2013) 659-672.
- [16] De Bock, K. W. and Van Den Poel, D. Reconciling performance and interpretability in customer churn prediction using ensemble learning based on generalized additive models. *Expert Systems with Applications* **39** (8) (2012) 6816-6826.

- [17] Karahoca, A. and Karahoca, D. GSM churn management by using fuzzy c-means clustering and adaptive neuro fuzzy inference system. *Expert Systems with Applications* **38** (3) (2011) 1814-1822.
- [18] Keramati, A., Jafari-Marandi, R., Aliannejadi, M., Ahmadian, I., Mozaffari, M., and Abbasi, U. Improved churn prediction in telecommunication industry using data mining techniques. *Applied Soft Computing* **24** (2014) 994-1012.
- [19] Keramati, A. and Ardabili, S. M. Churn analysis for an Iranian mobile operator. *Telecommunications Policy* **35** (4) (2011) 344-356.
- [20] Kisioglu, P. and Topcu, Y. I. Applying Bayesian Belief Network approach to customer churn analysis: A case study on the telecom industry of Turkey. *Expert Systems with Applications* **38** (6) (2011) 7151-7157.
- [21] Lu, N., Lin, H., Lu, J. and Zhang, G. A customer churn prediction model in telecom industry using boosting. *IEEE Transactions on Industrial Informatics* **10** (2) (2014) 1659-1665.
- [22] Owczarczuk, M. Churn models for prepaid customers in the cellular telecommunication industry using large data marts. *Expert Systems with Applications* **37** (6) (2010) 4710-4712.
- [23] Dalvi, P. K., Khandge, S. K., Deomore, A., Bankar, A. and Kanade, V. A. Analysis of customer churn prediction in telecom industry using decision trees and logistic regression. *Symposium on Colossal Data Analysis and Networking (CDAN)*, 2016, 1-4.
- [24] Tsai, C. F. and Lu, Y. H. Customer churn prediction by hybrid neural networks. *Expert Systems with Applications* **36** (10) (2009) 12547-12553.
- [25] Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G. and Chatzisavvas, K. C. A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory* **55** (2015) 1-9.
- [26] Verbeke, W., Martens, D., Mues, C. and Baesens, B. Building comprehensible customer churn prediction models with advanced rule induction techniques. *Expert systems with applications* **38** (3) (2011) 2354-2364.
- [27] Bi, W., Cai, M., Liu, M. and Li, G. A big data clustering algorithm for mitigating the risk of customer churn. *IEEE Transactions on Industrial Informatics* **12** (3) (2016) 1270-1281.
- [28] Zhao, L., Gao, Q., Dong, X., Dong, A. and Dong, X. K-local mamum margin feature extraction algorithm for churn prediction in telecom. *Cluster Computing* **20** (2) (2017) 1401-1409.
- [29] Datasource-<https://www.ibm.com/communities/analytics/watson-analytics-blog/guide-to-sample-datasets/>
- [30] Raja, J. B., Pandian, S. C. and Pamina, J. Certificate revocation mechanism in mobile ADHOC grid architecture. *Int. J. Comput. Sci. Trends Technol.*, **5** (2017) 125-130.
- [31] Lydia, E. L., Shankar, K., Pamina, J. and Raja, J. B. Correlating NoSQL Databases with a Relational Database: Performance and Space. *International Journal of Pure and Applied Mathematics* **118** (7) (2016) 235-244.
- [32] Raja, J. B. and Rabinson, K. V. Iaas for Private and Public Cloud using Openstack. *International Journal of Engineering Research* **5** (04) (2016).