Intelligent Big Data Analysis to Design Smart Predictor for Customer Churn in Telecommunication Industry

Samaher Al_Janabi(✉) and Fatma Razaq

Department of Computer Science, Faculty of Science for Women (WSCI), University of Babylon, Babylon, Iraq
samaher@uobabylon.edu.iq, razaqfatima200@gmail.com

Abstract. Due to the extensive development in the field of telecommunications, so today, it requires that companies must be familiar with and understand the nature of customers and their aspirations. This has led to strong competition between these companies so that they need to use the programmers to build accurate analysis systems to maintain their customers and improve the level of revenue. Here we have introduced an integration system that helps the telecom company achieve that goal. The proposed system consists of three basic pashas: First Phase: An understanding of the company's data, which consisted of two main parts, included data on the same company in terms of number of employees, number of customers, revenues and expenses, and customer-related data. This phase focuses on initial processing of data that is fragmented and unbalanced. Where the data of the company and the customer was merged first using the joiner and then we addressed the problem of imbalance by building DSMOTE algorithm, which adopted the principle (samples and quadratic) and succeeded in the production of real samples instead of default in the treatment of imbalance. Second Phase: Data were separated after processing into training and testing data. The training data were used to construct a GBM-based predictor after it was developed and replace its decision-making part, which is (DT) with a (GA) algorithm to identify customers to three groups are (the group of customers most influencing the company's revenues, the group of medium-sized customers and the least important group of customers). Third Stage: The accuracy of the predictor results was verified by using the matrix of the conflict matrix which are: Accuracy, Precision, Recall, F_measure, Fb. A comparison was made between the traditional method of initial treatment, which is SMOTE, DSMOTE in terms of error rate and accuracy. The best results for the developed method were when the data was divided by (40:60) and the error rate was (0.038) and the correct rate (0.962) while the traditional method was the best results for error rate (0.198) and resolution rate (0.802). In addition, the results of the GBM and GBM-GA were compared in terms of the four contrast matrix scales. The traditional method of GBM had the value of Accuracy (0.88), while the developed GBM-GA method was Accuracy (0.97). This confirmed the accuracy of the proposed method.

Keywords: Intelligent data analysis · Customer churn prediction · Imbalance dataset problem · DSMOT · GBM-GA · Confusion matrix measures

Y. Farhaoui and L. Moussaid (Eds.): ICBDSDE 2018, SBD 53, pp. 246–272, 2019.
https://doi.org/10.1007/978-3-030-12048-1_26

1 Introduction

The problem of customer churn is classified into three main classes [14, 15, 18] which are voluntary churners, non-voluntary churners, and silent churners. Customers that would like to end their contracts in a company to participate in another one is called voluntary churners. There are a number of reasons that motivates the customer to churn in this situation. For example, one reason might be the availability of new technology supported by another company than is not yet supported by the current company he/she is subscribed to. Other reasons include expiration of the contract, degradation of the quality, and others. Non-voluntary churners are those that abuse the service provided by the company or not disbursing the bill on time and thus the service provider retract their service. As a result, such customers are forced to find another company. Silent churners are those that left the company without any prior reason that the company can make use for reducing the future churners.

Customers always seek for better quality with less price in Telecommunication sector. Thus, they make benefit from the offers provided for new customers and after the end of such offers, they can leave the company without any restrictions to find another company [19]. Companies do their best in order to encourage customers to move to them [20].

Because new customers are likely to leave the company, they should not be considered in the churn rate estimation. But losses of such new customers in their subscription period could be or not to be considered. When included, the churn rate will include the new customers who left the company. Otherwise, it measures how many of the initial customers have left the company. The churn rate could be calculated in several ways such as:

- Use a fixed period of time such as a month or a year.
- Get the number of customers that left the company within a period
- Split this quantity by the number of customers that the company have at the beginning of this period

2 Literature View

There are many researches work to solve the customer churn problem and they used multi techniques to deal with this problem as explain below:

Zhu and Baesens [6], presented a novel customer churn model for prediction using a mobile application, this predictor pre-processing standardization by logarithm transformation to obtain the best results. This approach deal with data stream clustering (i.e., the aim in this step, to understand the data stream, and find out the customers distribution) then customer churn prediction (i.e., the aim to predict the probability of a given customer abandoning his or her smartphone in the future, at the prediction phase, based on the definition of churn behaviours in the analysis phase), results show an accuracy rate of 87%. This paper deal with customer churn predication problem and it used a measure from the confusion matrix same as our work, while it different in pre-processing and predication method.

Machado et al. [7], a shown an approach to enhanced feature mining and classifier models to predict customer churn for an e-retailer, the pre-processing used filter approach's which are algorithms used to rank the prominent features influencing customer churn and making sure the noisy features with no relevance are ignored baseline methods filtering the feature-set and imbalance in output class labels, they used gradient boost ensemble classifier, SVM Classifier, logistic regression with L1 to determine performance and calculated the validity criteria, they demonstrate how regularized logistic regression, based on comprehensive cross-validation techniques, it is proved that support vector machine and gradient boosting algorithm using random forests are the best models for working with the problem of customer churn. This work suggests methods to handle customer churn predication problem and it used cross validation principle and in these two points similar our work, but differ types of the methods that used.

Subramanya and Soman [8] explained an approach about customer churn prediction that uses Strong Social Ties, dyad churn model using strong social ties (i.e., (i) mobile call graph is constructed using call detail records collected over a predefined period of time, (ii) the strength of social ties among the constructed graph nodes is evaluated using a new set of proposed social attributes, (iii) the strongest ties within a specific percentage are kept in the call graph, (iv) an influence propagation model is applied and the set of influenced nodes are determined to represent the set of future potential churners). The results also showed that strong social ties were among the effective ways of determining the churn compared to weak ties. Similarity in predicting the customer churn and "length of calls" between customers is the most effective attribute in the predictions. This paper differs in our work by it presented a novel method for evaluating social tie strength between Telecommunication customers, and introduced new attributes, while prediction accuracy measured by a lift chart.

Abd-allah [9] ppresented an approach to customer lifetime value (CLTV) prediction, the state of the art in this domain uses large numbers of features and ensemble regressors to forecast value, the approach is embedding (solid arrows represent the flow of data), and the dashed arrows represents interaction between stakeholders and systems/data, it used random forests for churn classification and CLTV regression, the resulting prediction are piped to operational systems, they obtain spearman rank-order correlation coefficients of 0.56 (for all customers) and 0.46 (excluding customers with a CLTV of 0). This paper differ on our work by using system Microsoft Azure blob storage (ASOS) for provides daily estimates of future delay of every customer and it used two pre-processing stages (i.e., split data using ASOS after aggregation and pipeline).

Chamberlain et al. [10] suggested an approach for early prediction of customer churn with personalized targeting in social games in mobile devices. A binary classification churn algorithm for predicting the likelihood of predicting the churn of a customer based on using behavioural data from the first day of lifetime (i.e., using several algorithms Gradient Boosting, Decision Tree, Random Forest, Logistic Regression, and Gaussian Naive Bayes) and churn prevention that is based on push notifications (either containing their favourite feature they not explored yet) able to reduce churn up to 28%, which, at the scale of millions of users. The similarity in this paper with our work in analysing the results obtained from performing 10-fold

cross-validation on our machine learning model, while the different in pre-processing and using for performance evaluation several metrics (the model is evaluated using F-1 score and AUC (area under curve).

Milosevic et al. [11] presented an approach for telco churn prediction: building the model, evaluation and network architecture extract network features in a process called featurization, the approach is social network analytics (SNA) by using relational learners (RL) is apply on graph customers, and after the RL has been applied. Each customer now has a score or probability of churning and then to building churn models for predicting the likelihood of churn using non-relational classifiers with network features and scores from relational learners, results in p-values of less than 0.01 for all performance measures the non-relational classifiers with different set of features are great compared to relational learners. This paper similar our work in handle the same problem (i.e., customer churn prediction) and preprocessing dataset before build model, while, different in used SNA in churn prediction and it used Performance measures (lift, AUC, EMP).

Long et al. [12] presented an approach for churn prediction in Telecommunication, the data pre-processing steps include removing unwanted features, filling missing values, the normalization and discretization, the approach is a feature extraction algorithm using K-local maximum margin (i) Building a generalized data for each category for estimating the anisotropy factor σ, (ii) Euclidean distance used to select the nearest k points; (iii) Using the initial feature weight vector (w) to obtain the initial local hyper plane expansion coefficients of each sample combined with the gradient descent method to calculate and update the feature weight vector (w), (iv) K points are used to divide the corresponding hyper plane, this paper handle the same problem that our work attempt to solve it (i.e., churn predication industrial tetech) but differ on it in pre-processing, methodology, and evaluation measures.

Zhao et al. [13] focus on an approach for Intelligent churn prediction for Telecommunication using GP-AdaBoost learning and PSO under sampling, the pre-processing by input feature anomalies such as missing or empty features, some filters are responsible for converting the nominal features in the dataset to numerical format. This is by grouping the instances in different categories with different sizes, Ch-GPAB approach works by evolving multiple GP programs for each class using the AdaBoost technique, based on the results it is proved that the proposed system yields 0.91 AUC and 0.86 AUC on Cell2Cell and Orange. This paper similar with our work in two points (i.e., handle the same and handle the imbalance data but differ in the techniques that used and evaluation measures).

Idris et al. [14] shown an approach for class imbalance brings challenges to the problem of predicting the customer churn, preprocessing in tow step (i) dealing with missing value (ii) feature selection approach, the approach used four ensemble methods (i.e., there are 4 main types of ensemble solutions which are bagging, boosting, random-forest and hybrid), while experimental results show that the used metric has impacts on the performance of techniques greatly, this paper similar with our work in handle imbalance problem but it differ in used four ensemble methods and performance evaluation measures.

Vijaya et al. [15] suggested an efficient system for predicting the customer churn which employs particle swarm optimization (PSO) and proposes three variants of PSO for predicting the churn which are PSO with feature selection as a pre-processing step, PSO with simulated annealing (SA) embedded with, and finally PSO with different ways for feature selection and SA, Result with three cases Performance on Orange-1000 and Performance on Orange-5000 and Performance on Orange-7344 and found the performance has improved for the Orange 7344 compared to the Orange 5000 and Orange 1000. This paper with our work in used algorithm to for churn prediction and handle imbalance while, differ in technique used use and set of evaluation metric.

3 Main Concepts of the Task Problem

In this section, we will show the main concept used to definition and solve problem.

3.1 Intelligent Data Analysis (IDA)

IDA is considered one of the key areas in real applications and computer science [2, 17]. The concept of intelligent data analysis to develop new ways of indicating during the discovery or recovery pattern preparing us learn the tools to find patterns of data. The objective of the analysis is the purpose of building rules or troubleshooting instabilmente optimization or classified data or prediction of values or make a summarized in smart and understandable manner [1]. Figure 1 explain, how can recruit IDA to solve the problem under hand.

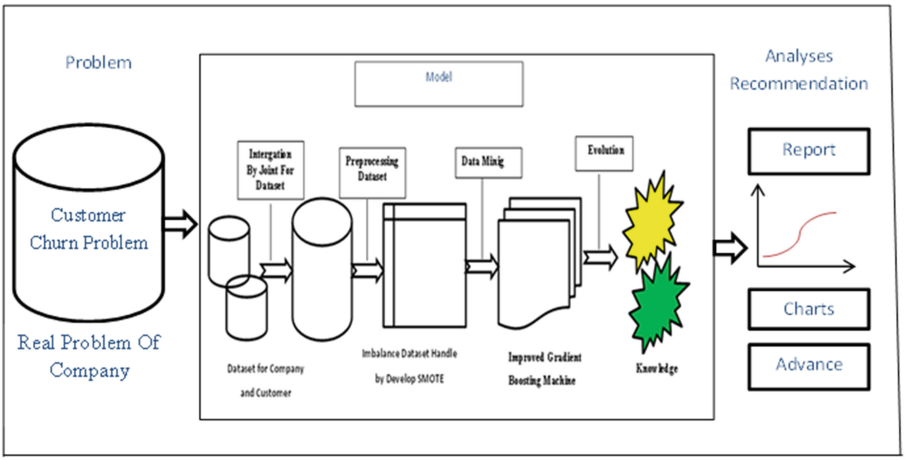


Fig. 1. IDA process stages

3.2 Imbalance Dataset

A data set is called imbalanced if it contains many more samples from one class than from the rest of the classes. A dataset is called to be unbalanced when one class has more samples in the training than another. This will result a classifier that has good performance with the majority class that has many training samples but it will have poor accuracy with the other class that has less training samples. Applications that has such problems are such as fault detection, medical diagnosis, and fraud detection [2, 5]. The solutions to the above problem fall in three main categories [5]:

- Data-level
- Algorithm-level
- Ensemble solutions

Data-level solutions resamples the data as a pre-processing step to help reduce the imbalance in the dataset. Algorithm-level solutions not work with the data as its name suggest but modify the algorithms or create new ones to make the learning process more biased towards the minority class. Ensemble solutions either modify the ensemble learning algorithms at the data-level to pre-process the data or by embedding a framework that is cost-sensitive within the ensemble learning process. Each solution of the three above solutions has its pros and cons. The major issue is that the imbalance levels in churn data cannot be determined earlier. It changes to a great extent between organizations. Thus, churn detection systems must be able of recognizing the imbalance levels and apply suitable balancing methods on the data such that the classifier is adequately trained across all data classes [5].

3.3 Prediction Techniques

Prediction tries to solve a variety of problems by producing a set of rules that helps reaching a reasonable solution for a specific problem. In this section, the major properties of four prediction algorithms have been considered and a comparison among [3].

- **Classification and Regression Tree (CART)**

CART is one of the decision trees techniques used to classify data in an easy way. Based on input variables (X), the target variable (Y) is predicted. By recursively splitting the data from top to bottom, the decision tree is to be built. Each branch of the tree represents different values for the input variables (X). After looking at all values for such variables, the decision is made at the terminal nodes at the end of the tree. In each split, just one variable is used [6, 14]. As a result, too many levels will be needed if many variables exist and thus more computational time [3].

- **Multivariate Adaptive Regression Splines (MARS)**

MARS uses the "divide and conquer" principle for building the model. It uses multiple equations where each one represents a region in the input space. This approach can handle multi-dimensional data [21]. Because it does not have a tree as in CART, it depends totally on the mathematical functions in finding the optimum solution. Despite being fast in predictions, its accuracy could be affected by the discontinuity in the boundaries between the different regions [22].

- **Support Vector Machine (SVM)**

SVM is a classification approach that uses the principle of decision boundary to determine the class value. Because linearly separable problems are simpler, SVM mitigates the binary class problem by applying linear separable input space and also converts the non-linear separable input space into a another linear space. This concept allows locate the optimal line of decision with the maximum margin to resolve the issue of overfitting [3, 23]. It requires preprocessing time for transforming the variables from a non-linear space to a linear space. Parameters existing are the kernel function and the cost function.

- **Gradient Boosting Machine (GBM)**

GBM creates a strong learner by using a number of weak learners. It converts the problem into a regression problem and tries to reduce the prediction error as much as possible by aggregating multiple weak learners. The typically used weak learner is the decision trees but any other weak learner could be used. GBM can also be reduced to regression with a loss of function to address classification problems [3]. GBM algorithm builds an additive model for minimizing residual which represents the difference between the predicted and target values [24]. The target value for each data record is re-estimated at each iteration until the least possible error is reached [10]. The linear combination of many binary regression trees represent the final model of GBM. The results of GBM usually are much more stable [10]. The parameters that have a major effect on the performance and behaviour of GBM are as follows [4, 16, 25]:

- The first parameter is the maximum number of decision trees algorithms used to create the GBM model. This is because the larger the number of algorithms to use the more complexity the model will be and thus it will be time-consuming.
- The second parameter for GBM is the learning rate or shrinkage. It helps to reduce overfitting which may happen when fitting the data in training process too well but it has poor prediction accuracy for future unseen samples. Small value of the learning rate is preferred to train the model where the best result corresponds to using a learning rate less than 0.1. Reducing the learning rate value might increase the number of trees used.

- Maximum number of terminal nodes is the third parameter to be used to avoid overtting. Usually, all decision trees have an equal number of terminal nodes. The smaller number of terminal nodes, the simper the algorithm and thus less computational time required for its operation. It allows us to understand and interpret it.
- The fourth parameter of GBM is minimum number of data records in a terminal node that helps to reach the stopping condition for the splitting process. When the number of records is less than a particular value, the splitting will stop. Small number of this parameter produce a large BRT and vice versa.

3.4 Churn Prediction

Churn predication is give data around which customer are most likely to take off the benefit in the close future, there are numerous distinctive reasons for customer to churn like move domestic, relentless, others like sudden death and undetectable. The center is put here fair on churn expectations or maybe churn and non-churn expectations as only churn predictions are significant and ordinarily cause a few costs. There is no activity and no costs included in a forecast of a non-churner but the rate of wrong positives forecasts, that is non-churners that were inaccurately classified as churners, limits the greatest number of churners that can be picked [1].

Table 1. Comparison among main techniques of predication

Prediction techniques	Disadvantages	Advantages
CART	• Don't have unstable DTs • Splits only one variable at a time	• Automatic class balancing • Handles the missing values automatically • Easily handle outliers with justification
MARS	• Discontinuity in sub-region boundaries that enhance accuracy • Needs backward steps to fix over fitting	• No user specific parameters • No variable transformation • More flexible • Inappropriate variable selection • Fast predication but error prone
SVM	• Superior variables transformation • Uses two user specific parameters	• Global minimum objective function to reach the best accuracy • Maximizing the margin to control capacity also increases accuracy • Able to work with categorical data with dummy variables which increases the accuracy
GBM	• Requires all data to be available before training the model. This does not fit for large datasets which could not fit once in memory • The selection of values for the four parameters is critical in the successful build of the GBM	• High accuracy due to fixing the errors of previous models with new ones

Customer churn is the term utilized by a variety of commerce to denote the movement of a customer from a benefit provider to another. There are different reasons for this development, such as benefit cost and quality, and loyalty to a benefit provider is specifically related to those criteria. In hone, the cost of attracting an unused customer is 5 to 6 higher than to keeping an as of now loyal customer. As a result, companies are always underweighted to reduce customer churn [1].

Predicting the customer churn is especially imperative and has as of late received more consideration from a variety of businesses such Telecommunication operators and as smartphone producers. Advances in these businesses such as modern services and technologies, as well as advances in the regions of machine learning, data mining increased the competition in the market [6].

4 Smart Customer Predictor (SCP)

The basic idea of any bussnices companies are their customers, so, each company tries to keep its own customers, and to win new customers from the surroundings, to increasing its revenue by improving the company's services and providing support for the companies. The main problems; tries these faces the companies are as follows (a) Leaving customers the large companies to a branch companies that has recently opened, which give the customers better offers and services. (b) The company services may become bad, resulting in, the client will leave the company to find a better service. (c) The customers sometimes leave the services of these companies for a period of time, and then leave the services of this company either to leave the country or for other reasons.

These data are divided into two parts separate from each other, first part for the customer (customer communications, the messages sent by the customer, number of times the customer is charged). And second part for the company includes (company revenue, number of employees in the company, type of services provided by the company). Therefore, the problems of these data are:

- It needs for integration to connect between these two parts
- It needs to preprocessing stage, when we reviewing the databases of all telecommunications companies, we found it suffer from a problem of imbalance data, which mean the number of the customers that leaving the company is equal or more to the number of customers that remaining in the company, so, our goal is how to find a preprocessing method to solve this problem. After the data is merged and by one of the processing algorithm of imbalance which called "smote" that proven successes in processing the imbalance, we found this algorithm depend on coefficients are the number of the neighborhoods, and duplicate the amount of data, which mean the original data which multiplies by a hundred percent 100%, 200%, 300%, 400%, 500% when multiplied 100%.

The number of neighborhoods is one this method not compatible with prediction principle, because the prediction will be not correct if the data was virtual so it cannot merge the smote in it's current way with the prediction that has been built, so the main goal was how to improve the "smote" and make it built real data (not virtual) by depending on two coefficients which are using the principle of 10-sampling each time divide it into unequal proportions Instead of doubling the data and built virtual data also the number of the neighborhoods was depending on the number of duplicate, we found the best way to determine the number of the neighborhoods is by using the principle of the "quartets", and after that, evaluate the result of this step and identify the best division the number of neighborhoods, after that, built the prediction by one of the data mining algorithms which give the good result.

- Building model, Prediction techniques are an effective tool for discovering the knowledge from large and complex databases of the customers. In this work, a predictive model is designed and implemented that analyzes customer data and provides sufficient information to prediction customers leaving the company. The proposed model, it includes building predication model using improved gradient boosting machine by replace the core of it is (Decision Tree) by (genetic algorithm). We use main measure of predication error to evaluate the final model, these measures include confusion matrix: Accuracy (AC), recall or true positive rate (TP), precision (P), F-measure (considers both precision and recall) and Fb and 10-fold cross validation.

Algorthim#1: SCP
Input: *: Databases have contract and behavior dataset*
Output: *Build of SCP model*
- ***Set:*** *F1:First File, F2:Scound File,F3:Thrid File ,K1:List represent of neighbors quadratics , T1:List of number of minatory class samples, NF: number of attributer,T1:Array for minatory class samples, S: Array for new synthetic samples, CT : Array contains cases training data set based on 10-Sampling, F: array of training dataset, Original _target: array for Original _target T:traget,* $t_{1,\ldots,j}$*: target value, M: the mean of targeted values, NDR: number of data rows, DT: Training Data, LR: Leaner Rate, values real, values predication.*

Step1*: preprocessing dataset*
Step1.1:*F3= Call Joiner (F1,F2)*
Step1.2: *S[new][attrib]= Call DSMOTE(F3)*
Step 2: *Build Smart Customer Predictor*
Step2.1: *Split dataset according 10-fold Cross Validation into training and testing_dataset*
Step 2.2: *For each Training part not used*
- *Call DGBM-GA*
- *Test stopping condition satisfy Go to Step 3*
- *Else Go to 4*

Step3: *Evaluation and analysis behaviors of SCP*
- *Call Evaluation DGBM-GA*

End Pseud code

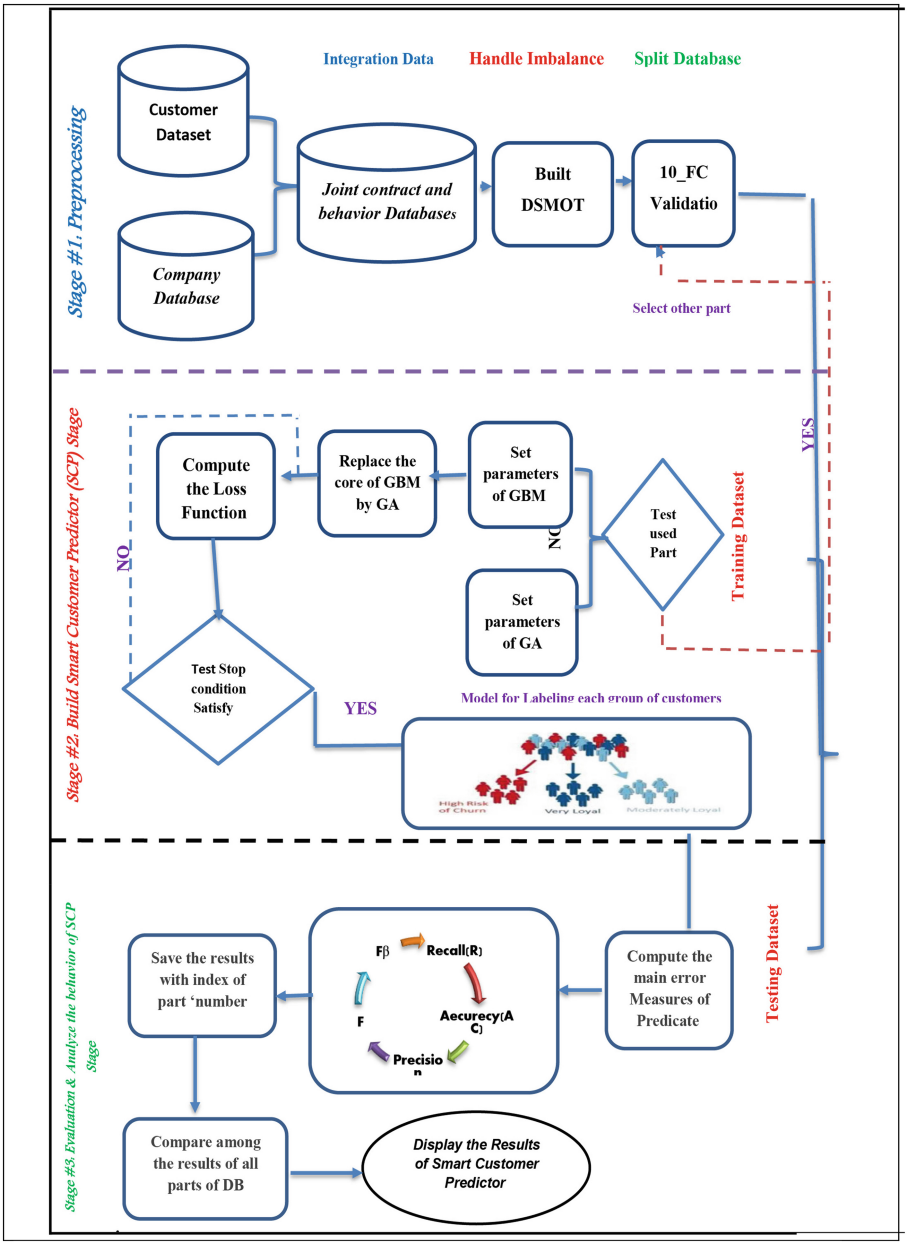


Fig. 2. Block diagram of SCP

4.1 Main Stages of Design (SCP)

In this section, building effectiveness predication model include three stage as explained; the first stage is dataset preprocessing that includes data collection, join the contract data and the behavioral data, apply DSMOTE algorithm be used to handle imbalance problem and create useful information help to building the smart churn prediction (SCP), The second stage, replace the core of Gradient Boosting Machine "i.e., Decision Tree" by Genetic Algorithm (GA) to create new predictor called (SCP). The final stage is evaluation the results based on confusion matrix measure and 10 cross validation.

A. Integration Stage

It is the first stage, after data is collected. In this work, we deal with joiner as example of integration data.

- **Joiner**

Database Joiner is connected data from an external table to incoming data before preprocessing began. The Database joiner queries an external table to retrieve contract data associated with behavior data. One or more columns exit in contract dataset are joined to one or more columns in behavior database the Algorithm 1 explained joiner steps.

Algorithm #3.1 : Joiner

Input: *Datasets(i.e., two files from different dataset)*

Output: *Joiner the Two File*

Set: *F1:First File, F2:Scound File,F3:Thrid File, diff_index: array for keeping index of columns not repeated*

Bool=counter for add different index.

Step1: *Read first file and second file*

Step2: *Compare header of two files*

Step2.1: *Fetch content header of two file*

Step 2.2: *Compare header length of first file with header length of second file*

- *If header_f1.length>=header_f2.length*
- *For int i=0 To header_f1.length*
- *For int j=0 To header_f2.length*
- *If header_f1[i]. equals header_f2[j]*
- *Bool=1*
- *If Bool==0*
- *Diff_index.add=i*

Step 2.3: *Repeat step2.2 when header of second file is larger*

Step 3: *Generated array of merge content the two file in single file*

Step 3.1 : *If content_f1.size()<content_f2.size()*

- *For int i=0 To content_f1.size()*
- *Line1=fetch of first file feature*
- *Line2=fetch of second file feature*
- *from line2, some data is required*
- *For int j=0 To diff_index.size()*
- *concoct the two array line 1 and line2 in line1*
- *index=i*

Step 3.2: *now complete from the second file*

- *For int i=index To content_f2.size*
- *For int j=0 To header_f1.length*
- *line_1[j]=" "*
- *from line2, some data is required*
- *for int k=0 To diff_index.size()*
- *Joint the two array line 1 and line2 in line1*

Step 3.3 *Repeat step (3.1,3.2, and 3.3) when , If content_f2.size<content_f1.size*

B. Preprocessing Stage

After completed connected the contract dataset with the behavioral dataset, through the joiner step, the result database may suffer from many other problems therefor, we will perform preprocessing on it.

- **Imbalance Problem**

Imbalance problem is a common problem in many applied data science and machine learning problems, a data set is called imbalanced if it contains little samples from one group than from the rest of the groups. Data sets are not balanced when at least one group is represented by only a small number of training examples (called the minority sampling) while another majority sampling make up the majority. In thesis, we will handle the imbalance problem of customer churn in telecommunication companies by Develop Synthetic Minority Oversampling Technique. The traditional SMOTE algorithm was suggested to face the impact of having a little representative of the minority group of samples in a database. Samples are taken from the minority groups and calculate the closest neighbors depends on the amount of oversampling required. The neighbors are randomly selected using the k nearest neighbors, (e.g. we select five neighbors randomly, based on the needed amount of over-sampling (e.g. we will select three neighbors out of five, if the oversampling value is 300% and create a single sample toward each neighbor. The Algorithm 3.2 explained SMOTE steps, we will illustrate by example smote percentage, when the dataset contains 784 row there are 178 sample (24%) of churn1, and 570 sample (76%) of Churn0. Although this isn't terribly balanced, Churn 1 represents people who leave the company. To increase the number of samples and make them balanced, we use a multiplication of amount of samples by 100% or 200% as we will explain in Table 2.

Table 2. Percentage rate of dataset

	Churn 0	Churn 1	Overall dataset
Original dataset	570(76%)	178(24%)	748
Amount of sampling = 100%	570(62%)	356(38%)	926
Amount of sampling = 200%	570(52%)	534(48%)	1104
Amount of sampling = 300%	570(44%)	712(56%)	1282

- **Developed SMOTE Algorithm**

Traditional SMOTE algorithm handles the problem by generates neighbors depending on the number of oversampling samples. Such that if oversampling is 200%, this will lead to number of neighbours equal to 2. As a result, this leads to a generated virtual population. While the goal of prediction techniques in telecommunication industry is finding results with high quality and speed. The predication law says, ***we can generate true predicator value only when the predictor build base on real values otherwise, the predictor generated false values (virtual values).*** Therefore, the smote is not suitable tool to use with any predictor, because it is generated virtual values to handle imbalance

problem. At the same therefore, this paper aiming to develop SMOTE algorithm to satisfy this law. DSMOTE will be developed over SMOTE by using ten sampling instead of arbitrary percentage (e.g. 100%, 200%, etc.) to split the data. And the Quadratic Principle (e.g. 2^1, 2^2,… etc.) to detect the numbers of neighbours, instead of depending on the number of samples, The evaluation of DSMOTE is based on both error and correct rates. The Algorithm 3 Explained DSMOTE steps. We can summarize the steps of Developed SMOTE Algorithm. [Split dataset into training and test dataset, Apply DSMOT on training dataset, Evaluate the results based on two measures (i.e. error and correct rate) to determine: Best number of neighbors. Best split based on 10-Sampling. Finally, suitable values to handle imbalance problem].

Algorithm # 3: DSMOT

- ***Input:*** *N1:List represent collection databases to customer and company, K1:List represent of neighbors quadratics , T1:List of number of minatory class samples*
- ***Output:*** *S[new][attrib]: array of new synthetic*
- *Set: NF: number of attributer,T1:Array for minatory class samples, New: counter for save number of S generated, S: Array for new synthetic samples, CT : Array contains cases training data set based on 10-Sampling, N2: variable of keep random number between 1 and K1*

Step 1: *Split the selected database into Training and Testing Database using 10-sampling for Generated Minority Of Populate.*

- *Tindex=0*
- *For L=1 To 9*
- *For J=0 To CT. Length*
- *If CT. Length <>0 && CT[J]<T1*
- *T1=TC[J]*
- *Tindex=J*

Step2*: compute amount oversampling*

- *N1=(int)(N1/100).*

Step3: compute *neighbors based on quadratics* $[2^0,....2^i],i=[1,2,3,4,5,6]$.

Step4 : *generated new synthetic samples based on 10_sampling*

Step4.1: *Selected random number of neighbors quadratics .*

- $N2=Rand(Max_1$ --- $Min_i) + Min_i$
- $NN=2^{N2}$

Step4.2 : *build loop for number of attributer :*

- *For attrib=1 To NF*
- *Different=T[NN[nn]][attrib]-T[i[attrib]*
- *Gap=generated random number between[0,1]*
- *S[new][attrib]=T1[][attrib]+Gap*different.*
- *New++*
- *End for*

End of pseudocode

C. Develop core of Gradient Boosting

In this section, we will develop the traditional gradient boosting algorithm through replace their core DT by GA, because decision tree has some drawbacks while could be summarized as follows:

- Difficult to determine the best number of trees. Adding more trees to model will increase the computational time. This also may lead to overfitting.
- Also, the tree depth deeply affects the training process. This is a second cause for overfitting if the depth was high.
- Time consuming when used with thousands of features.
- Decision trees are harder to interpret compared to other simple models such as linear models.
- Decision trees parameters are statically defined by the machine learning engineer and there is no way for the learner to find it itself.

We have to take care that such learner is simple and not time consuming in its calculations to be a suitable option for use as the gradient boosting core. Based on such discussion, genetic algorithm seems a good option. Genetic algorithm is a random-based optimization algorithm that is able to produce high quality solutions by evolving bad solutions. Motivations to use genetic algorithm over decision trees are as follows:

- Genetic algorithm can optimize itself to produce better solutions while decision trees cannot do that.
- Genetic algorithm can produce multiple optimum solutions to the problem and not bounded to just one solution.
- Genetic algorithm is able to avoid local optimum solutions as it follows multiple paths in parallel.
- Genetic algorithm searches the complete space for the best solution that solves the problem compared to decision trees that cannot search the complete space. Decision trees applies the solution that is fed to it by the engineer. The Algorithm 4 Explained DGBM-GA steps.

Next is to know how the genetic algorithm will do predication or satisfy the grouping based on rang. Genetic algorithm is an optimization technique that can tell what the best solution to a given problem is. There must be some modifications to be applied in order to make the genetic algorithm able to do grouping based on rang. Otherwise, Predicting the user's formula by dividing the community into three groups according to the ranges. Thresholding the sigmoid output is done according to the following formula:

$$predict\ group = \begin{cases} Active\ Customers, sigmoid_{out} > [1 - 0.6] \\ Modelate\ Customers, sigmoid_{out} > [0.3 - 0.5] \\ Passive\ Customers, sigmoid_{out} \geq [0 - 0.2] \end{cases}$$

Table: Predict Grouping of Customers based on DGBM-GA corresponding

Fold 1	DGBM-GA	# Active customer
		# Modelate customer
		# Passive customer
Fold 2	DGBM-GA	# Active customer
		# Modelate customer
		# Passive customer
Fold 3	DGBM-GA	# Active customer
		# Modulate customer
		# Passive customer
Fold 4	DGBM-GA	# Active customer
		# Modelate customer
		# Passive customer
Fold 5	DGBM-GA	# Active customer
		# Modelate customer
		# Passive customer
Fold 6	DGBM-GA	# Active customer
		# Modelate customer
		# Passive customer
Fold 7	DGBM-GA	# Active customer
		# Modelate customer
		# Passive customer
Fold 8	DGBM-GA	# Active customer
		# Modelate customer
		# Passive customer
Fold 9	DGBM-GA	# Active customer
		# Modelate customer
		# Passive customer
Fold 10	DGBM-GA	# Active customer
		# Modelate customer
		# Passive customer

Agorthim # 4: DGBM-GA[4]

Input: : Collection Of Databases Related To Customer And Company

Output: *Build Of Churn Customer Predication Model*

Set: *F: array of predicted values of training dataset, C:Counter of records, C_{GA}: counter of GA, Org_Goal: The array of original Goal of Training Dataset, G:Goal, $g_{1,\ldots,j}$: values of Goal*

Step1: *Split the selected database into Training and Testing Dataset using 10-Fold Cross Validation techniques.*

Step2*: Build the Training Stages of SCP Model:*

Step3: *Find the initial prediction for all data records in training dataset by:*

- *Calculating the mean of targeted values (Mean (G)).*
- *While C>N*
  - *F [0, C] = Mean(G)*
  - *Org_Goal[C] = Training _Dataset[g,C]*
  - *C= C+ 1*
- *End While*

Step4 *While C_{GA} >= GA_MAX, build predictor GBM model by:*

- *C= 0*
- *While C> N, Update target values of Training dataset by:*
  - *Residual [C] = Org_Goal[C] – F[C_{GA-} 1, C]*
  - *Training _Dataset [g,C] = Residual [C]*
  - *C= C+ 1*
  - *End While*

- *Call building GA (Training dataset) and retrieve final model of GA.*

Step5 *: For each GA in GBM:*

- *While C _ GA > number of data rows in GA , update prediction values by:*
- *F [C_{GA}, Cr _ GA] = F[C_{GA}-1, C_GA] + (Sk× GApredicated_value)*
- *Increase counter of data rows in final GA : C _ GA= C _ GA + 1*
- *Increase counter of GA C_{GA} = C_{GA}+ 1*

Step6: *Return predictive array of prediction values F*

Algorithm # 5: GA

Input: *Database.*

Output: *The best or optimal groups*

Step 1: *Set the number of embedded images, Max Gen to max epoch allowed.*

Set: *Gen ← 1*

Do *For each chromosome in the population*
Randomly choose the ki seed
Distribute these seeds in the chromosome.
End for
Repeat
Do For each chromosome in the population
Compute fitness.
End for
Keep the best individual.
Select *subpopulation of parents by applying (Roulette Wheel Selection).*
Perform *crossover between parents by applying uniform crossover with probability pc.*
Perform *mutation on child with probability pm.*
Perform *replace old population with new population.*
Gen ← Gen+1(apply Elitism principle)
Until *(Gen >Max Gen)*
Return the best groups.
End.

D. Evaluation Stage

In this section, we will explain the evaluation of the results in two phases. The first stage (evaluation the results of preprocessing stage) handles imbalance of the dataset (traditional SMOTE and Developed SMOTE) based on Error and Correct Rate. The second stage: evaluation model stage by develop GBM algorithm based on the confusion matrix, these matrix uses several performance metrics (accuracy, recall, precision and f-measure). Algorithm 6 explained Evaluation the Performance of DGBM-GA.

Algorithm #6: Evaluation The Performance of DGBM-GA

Input: *values real, values predication*

Output: *performance metrics*

Step1:value initial for True Positive, True Negative , False Negative ,False Positive
True Positive=0, True Negative=0 , False Negative=0 ,False Positive=0) (
Step2: build confusion matrices
Step2.1: For i=0 to length (values predication)

- *If values real [i] == values predication [i] == 1:*
- *TP += 1*
- *If values predication [i] == 1 and values real[i] != values predication [i*
- *:[FP += 1*
- *If values real [i] == values predication [i] == 0:*
- *TN += 1*
- *If values predication [i] == 0 and values real[i] != values predication [i]*
- *FN += 1*

Step3: Return (True Positive, True Negative, False Negative, False Positive)values
Step4: compute Evaluation the Performance of DGBM-GA

- $Accuracy = TP + TN/(TP + FN + FP + TN)$
- $Precision = TP/(TP + FP)$
- $Racall = TP/(TP + FN)$
- $f_{measure} = 2 * Precision * Racall/(Precision + Racall)$
- $Fb = (1 + \beta^2).Precision * Racall/(\beta^2.Precision + Racall)$

End peso code

5 Experiments and Results

The dataset is available publicly, this data contains 3333 instances. When analyzing data relative to the column of the Churn we found the number of people remaining in the company 483 and, while Number of people leaving the company is 2850, As well as can be illustrated by the percentage of churn is 14.491%, while Non_Churn Percentage is 85.16%. The non-churns customers percentage was much larger than churns customers percentage in the selected dataset that can give a difficult time to churn.

Table 3 show the data base of contact and behavior contain 19 features, we not used all these features but remove two features (i.e., the state and number phone). While convert the code area and churn from integer into string (http://www.sgi.com/tech/mlc/db).

Table 3. Description of database

Dataset	Dataset desorption
Account length	Its value is real. The period in which the account was active
Area code	Categorical
Int'l plan	The international plan is operational (YES, NO)
Vail plan	Voice Mail plan is operational (YES, NO)
Vail message	Number of voicemail messages
Day mins	Overall minutes daily that utilize
Day calls	Overall daily phoning
Day charge	Overall day charge
Eve mins	Overall evening minutes
Eve calls	Overall evening calls
Eve charge	Overall evening charge
Night mins	Nighttime minutes
Night calls	Overall nighttime phoning
Night charge	Overall night-time
Intl mins	The international utilized minutes
Intl calls	Overall phoning internationally
Intl charge	Overall cost of the International
Custer calls	Number of customer service calls provided
Churn	Customer Churn (Target Variable yes = churn, no = not churned)

Table 4. Imbalance rate based on different rate

Sampling	Split dataset		Imbalance rate
	Taring rate	Testing rate	
Sample #1	90%	10%	0.131
Sample #2	80%	20%	0.140
Sample #3	70%	30%	0.134
Sample #4	60%	40%	0.136
Sample #5	50%	50%	0.120
Sample #6	40%	60%	0.126
Sample #7	30%	70%	0.130
Sample #8	20%	80%	0.140
Sample #9	10%	90%	0.135

Table 5. Joint for behavior and contact dataset

Number of row	Area code	Intl charge	Intl calls	Night charge	Night calls	Eve charge	Eve calls	Day charge	Day calls	CustServ calls	Intl mins	Night mins	Eve mins	Day mins	VMail plan	Int'l plan	Churn	Account length
1	415	2.7	3	11.01	91	16.78	99	45.07	110	1	10	244.7	197.4	265.1	1	0	0	128
2	415	3.7	3	11.45	103	16.62	103	27.47	123	1	13.7	254.4	195.5	161.6	1	0	0	107
	. .	. .	. .	. .	. .	. .	. .	. .	. .	. .	. .	. .	. .	. .	. .	. .	.. .	. .
1666	415	2.86	1	7.46	101	102	10.49	30.45	93	2	10.6	165.7	238.3	179.1	1	0	0	99
	. .	. .	. .	. .	. .		. .	. .	. .	. .	. .	. .	. .	. .	. .	. .	. .	. .
3333	510	1.35	1	6.26	137	84	13.57	36.35	2	2	5	139.2	159.6	213.8	0	1	0	184

- **Integration Stage**

In this section, we will explain the dataset after performance joint between tables. Table 5 explain the result of joint between the behavior and contact dataset that content 18 features and 3333 samples. The purpose of this stage is preparing the dataset to extraction useful information from it.

- **Develop SMOTE**

In this case study, DSMOTE used ten sampling and estimation the number of neighbors based on Quadratic principle, we will evaluate the results of DSMOTE based on two measures (i.e. Error Rate, Correct Rate and Correct).

Table 6. Error rate DSMOTE based on the 10-sampling and quadratic neighbors.

# Neighbors	Sample #1	Sample #2	Sample #3	Sample #4	Sample #5	Sample #6	Sample #7	Sample #8	Sample #9
2^0	**0.94**	**0.938**	**0.943**	**0.945**	**0.947**	**0.962**	0.952	0.943	0.933
2^1	0.953	0.943	0.952	0.94	0.95	0.95	0.938	0.952	0.934
2^2	0.929	0.933	0.926	0.933	0.943	0.955	0.945	0.95	0.936
2^3	0.936	0.926	0.933	0.938	0.95	0.945	0.938	0.95	0.926
2^4	0.946	0.915	0.945	0.955	0.936	0.962	0.94	0.952	0.943
2^5	0.924	0.934	0.936	0.95	0.94	0.95	0.927	0.952	0.94

The above table explain the correct rate of smote based on choose different values of quadratic neighbors and different split of dataset using ten sampling. The best value of error rate is (0.038). The database is divide into 40% for train and 60% for test and number of neighbor is 2^4.

The above table explain the correct rate of smote based on choose different values of quadratic neighbors and different split of dataset using ten sampling. The best value of error rate is (0.962). The database is dividing into 40% for train and 60% for test and number of neighbor is 2^4.

Table 7. Correct rate DSMOTE based on the 10-sampling and quadric neighbor

# Neighbors	Sample #1	Sample #2	Sample #3	Sample #4	Sample #5	Sample #6	Sample #7	Sample #8	Sample #9
2^0	0.060	0.062	0.057	0.055	0.053	0.038	0.0486	0.057	0.0674
2^1	0.047	0.057	0.048	0.060	0.0504	0.0504	0.06260	0.048	0.066
2^2	0.071	0.067	0.074	0.067	0.057	0.045	0.055	0.050	0.064
2^3	0.064	0.074	0.067	0.062	0.0504	0.055	0.062	0.050	0.074
2^4	0.054	0.085	0.055	0.045	0.064	0.038	0.060	0.048	0.057
2^5	0.076	0.066	0.064	0.050	0.608	0.0504	0.073	0.048	0.060

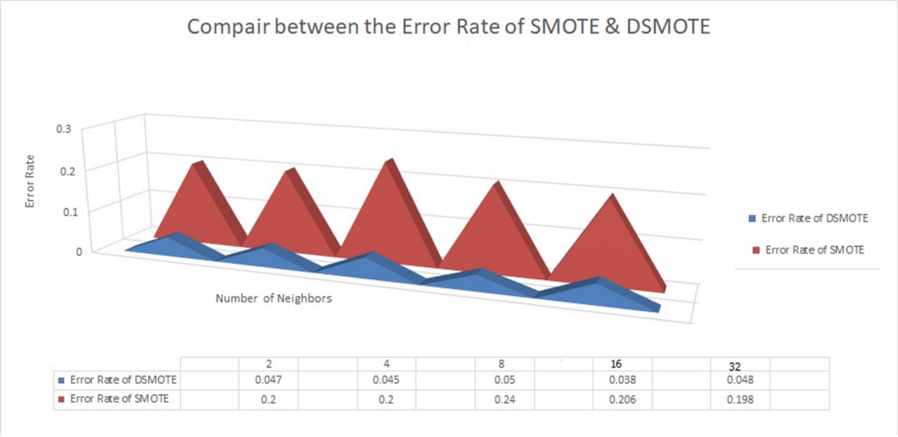


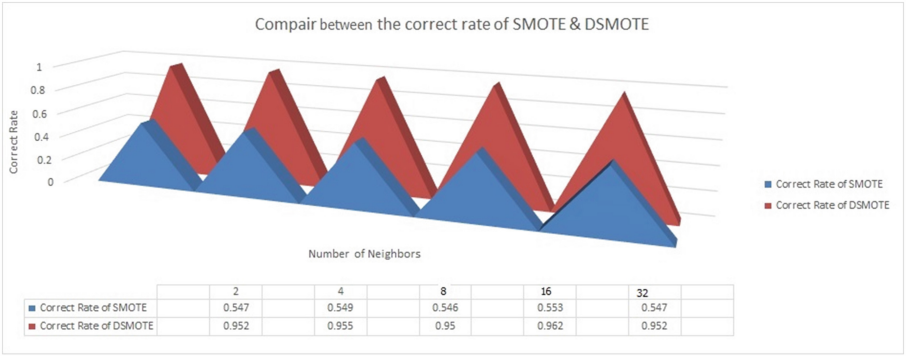
Fig. 3. Compare between the error rate of SMOTE and DSMOTE for best value for each test.

- **Apply GBM-GA**

To explained the benefit of built model SCP, compare with the Traditional GBM we need at the beginning determined the main parameter of it. As explained in chapter three in the Algorithm 4.

Choosing of parameters roles, the behavior of training process and affects the result of this process. Multiple values of Gradient Boosted Machine parameters have been used with Company dataset aiming to find optimal values for them.

Shrinkage parameter is tested with the rang of values (0.1–0.001) and the best result was with the range (0.005–0.008) with respect to another parameter.



	2	4	8	16	32
Correct Rate of SMOTE	0.547	0.549	0.546	0.553	0.547
Correct Rate of DSMOTE	0.952	0.955	0.95	0.962	0.952

Fig. 4. Compare between the correct rate of SMOTE & DSMOTE for best value for each test

Another parameter should be selected for GBM is the maximum number of terminal nodes which related with complexity of trees. The popular depth of regression tree in GBM is between 2 and 4, which means the maximum number of terminal nodes is in the range of (4–16) nodes. In GBM-GA, we not need for that parameter, while replaced by the main parameters related of GA that including (i.e., population size = 30, PC = 0.9, PMM = 0.001, Max number of generation = 50), while in traditional GBM depth of four levels is give stable results, in general, when we was using four levels which means the maximum number of terminal nodes with company dataset was (16) nodes that consider medium complexity for dataset with (19) attributes.

Third parameter in GBM is the minimum number of samples in terminal nodes. We also not need to determine this parameter in design GBM-GA.

In GBM the number of trees is the most important parameter and it must need to be chosen carefully, while, we not need to determine this parameter in design GBM-GA. In general, in traditional GBM used 100 trees.

Table 8 Comparison of original GBM and develop GBM-GA Based on Mean absolute Error

$$\text{Mean absolute error MAE} = \frac{1}{n}\sum_{t=1}^{n}|e_t|$$

Table 8 compares the mean absolute error when using decision trees and genetic algorithm as the core of gradient boosting for the training data. It is obvious that GA is better than DT from the first fold. Genetic algorithm seems to have its own mechanism to get out of local minimum that reduces the prediction error. But decision trees do not have such feature as it could not search the complete space as genetic algorithm does.

Table 8. Comparing results of GBM with (DT and GA) base on MAE.

			Fold #0	Fold #1	Fold #2	Fold#3	Fold #4	Fold #5	Fold #6	Fold #7	Fold #8	Fold #9
GBM	DT	MAE	0.320	0.271	0.132	0.960	0.768	0.582	0.453	0.380	0.339	0.812
	GA		0.238	0.200	0.130	0.436	0.832	0.307	0.151	0.860	0.162	0.560

Table 9. Compare the traditional GBM with develop GBM-GA

Performance metrics	GBM-GA values	GBM
Accuracy	0.97	0.88
Precision	0.79	0.37
Recall	1.0	0.126
F1	0.88	0.18

- **Evaluation**

To explained the benefit of evaluation stage. We will explain the results of both methods (Traditional GBM and develop GBM-GA) as explain in Table 9.

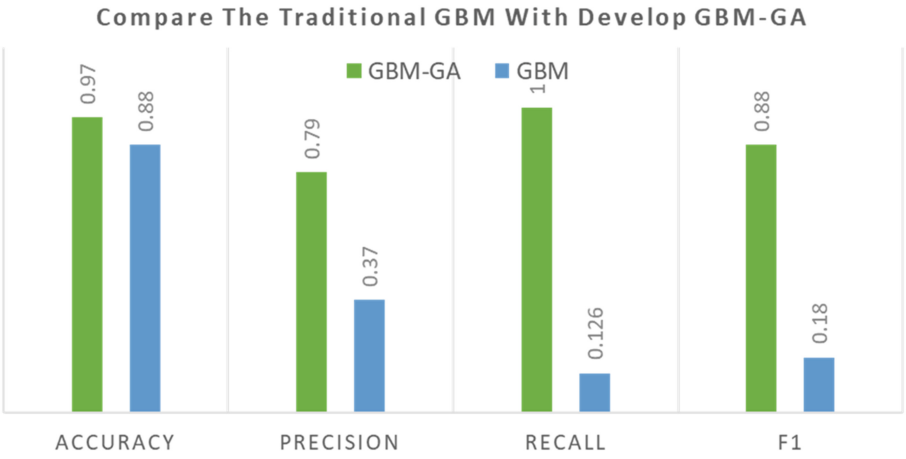


Fig. 5. Compare the traditional GBM with develop GBM-GA

6 Conclusion and Future Work

We can summarization the main points exclude from this paper by the following: Through the joiner the contract data and the behavioral data, we get Complete dataset to building the predictor, but this dataset remained suffer from the imbalance problem. DSMOTE technique is a powerful tool for solving imbalance problems. Compare with smote because it's held the advantage among SMOTE since it's not randomization. As explained in experiment more powerful and faster. Therefore, DSMOTE is consider method to use with prediction model because it led to true prediction. The impact of replace the core of Gradient Boosting Machine "i.e., Decision Tree" by genetic Algorithm (GA) is able to overcome DT problems, and reduce time implementation. The data that has been dealt with are relatively complex and composed of more than one part, so it was very difficult to build an accurate predictor achieves goal.

Depending on the customer segmentation, since this data in its abstract form does not give a precise perception for the company (e.g., number of calls, one person may make one call may take longer and more time than someone else who is making calls very little time). The Knowledge discover by the SCP determine the most important customers and influential on the company's revenue. A comparison was made between the traditional method of pre-processing, which is SMOTE, DSMOTE in terms of error rate and correct rate. The best results for the developed method were when the data was divided by (40:60) and the error rate was (0.038) and the correct rate (0.962) while the traditional method was the best results for error rate (0.198) and correct rate (0.802). In addition, the results of the GBM and GBM-GA were compared in terms of the four contrast matrix scales. The traditional method of GBM had the value of Accuracy (0.88), while the developed GBM-GA method was Accuracy (0.97). This confirmed the accuracy of the proposed method the concept of clustering for the remaining customers has been achieved accurately, where the threshold principle has been used for the customers clustering. It has proved its efficiency in determining the type of customers and their impact on the revenues of the company. Using the range principle inside of the GBM corresponds with predictive principle, which usually produces values of a continuous type.

To the Future Work, we suggest using proposed DSMOTE to address the imbalance in the field other than the field of business. The concept of correlation or gain information ratio can be applied on database after the joint step to determine the most important feature. We suggest using the MARS algorithm that is absolutely dependent on the mathematical aspect as a predictor rather than GBM, while, we know that mathematical methods often lead to more accurate results. The GBM core can be replaced with genetic program, because GP Works the same principle the decision trees, which is to build population in the form of decision trees.

References

1. Wang, H.F. (ed.): Intelligent Data Analysis: Developing New Methodologies Through Pattern Discovery and Recovery. IGI Global, Hershey (2009)
2. Al_Janabi, S.: Smart system to create an optimal higher education environment using IDA and IOTs. Int. J. Comput. Appl. (2018). https://doi.org/10.1080/1206212X.2018.1512460
3. Vijaya, J., Sivasankar, E.: An efficient system for customer churn prediction through particle swarm optimization based feature selection model with simulated annealing. Cluster Comput. 1–12 (2017). https://doi.org/10.1007/s10586-017-1172-1
4. Al_Janabi, S.: A novel agent-DKGBM predictor for business intelligence and analytics toward enterprise data discovery abstract. J. Babylon Univ./Pure Appl. Sci. **23**(2), 482–507 (2015)
5. Jiawei, H., Pei, J., Micheline, K.: Data Mining: Concepts and Techniques, 3rd edn. Elsevier (2011). ISBN 978-0-12-381479-1
6. Zhu, B., Baesens, B., van den Broucke, S.K.L.M.: An empirical comparison of techniques for the class imbalance problem in churn prediction. Inf. Sci. **408**, 84–99 (2017)
7. Machado, N.L.R., Ruiz, D.D.A.: Customer: a novel customer churn prediction method based on mobile application usage. In: 13th IEEE International Wireless Communications & Mobile Computing Conference, IWCMC 2017, pp. 2146–2151 (2017)

8. Subramanya, K.B., Somani, A.: Enhanced feature mining and classifier models to predict customer churn for an E-retailer. In: Proceedings of the 7th International Conference Confluence. 2017 Cloud Computing, Data Science & Engineering, pp. 531–536 (2017)
9. Abd-allah, M.N.: DyadChurn: customer churn prediction using strong social ties, pp. 1–11 (2017)
10. Chamberlain, B.P., Liu, C.H.B., Pagliari, R., Deisenroth, M.P.: Customer lifetime value prediction using embeddings, pp. 1753–1762. Elsevier (2017)
11. Milosevic, M., Zivi, N., Andjelkovi, I.: Early churn prediction with personalized targeting in mobile social games. Expert Syst. Appl. **83**, 326–332 (2017)
12. Óskarsdóttir, M., Bravo, C., Verbeke, W., Sarraute, C., Baesens, B., Vanthienen, J.: Social network analytics for churn prediction in telco: model building, evaluation and network architecture. Expert Syst. Appl. **85**, 204–220 (2017)
13. Zhao, L., Gao, Q., Dong, X., Dong, A., Dong, X.: K- local maximum margin feature extraction algorithm for churn prediction in telecom. Cluster Comput. **20**(2), 1401–1409 (2017)
14. Idris, A., Iftikhar, A., Rehman, Z.U.: Intelligent churn prediction for telecom using GP-AdaBoost learning and PSO undersampling. Cluster Comput. 1–15 (2017). https://doi.org/10.1007/s10586-017-1154-3
15. Vijaya Saradhi, V., Palshikar, G.: Employee churn prediction. Expert Syst. Appl. **38**(3), 1999–2006 (2011)
16. Ali, S.H.: Miner for OACCR: case of medical data analysis in knowledge discovery. In: 2012 6th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT), IEEE, Sousse, pp. 962–975 (2012). https://doi.org/10.1109/SETIT.2012.6482043
17. Al-Janabi, S.: Pragmatic miner to risk analysis for intrusion detection (PMRA-ID). In: Mohamed, A., Berry, M., Yap, B. (eds.) Soft Computing in Data Science. SCDS 2017. CCIS, vol. 788, pp. 263–277. Springer, Singapore (2017). https://doi.org/10.1007/978-981-10-7242-0_23
18. AlOmari, D., Hassan, M.: Predicting telecommunication customer churn using data mining techniques. In: International Conference on Internet and Distributed Computing Systems, pp. 167–178. Springer, Cham (2016)
19. Coussement, K., Van den Poel, D.: Improving customer attrition prediction by integrating emotions from client/company interaction emails and evaluating multiple classifiers. Expert Syst. Appl. **36**, 6127–6134 (2013)
20. Owczarczuk, M.: Churn models for prepaid customers in the cellular telecommunication industry using large data marts. Expert Syst. Appl. **37**(6), 4710–4712 (2010)
21. Mansiaux, Y., Carrat, F.: Detection of independent associations in a large epidemiologic dataset: a comparison of random forests, boosted regression trees, conventional and penalized logistic regression for identifying independent factors associated with H1N1pdm influenza infections. BMC Med. Res. Methodol. **14**(1), 99 (2014)
22. Trevor, H., Robert, T., Jerome, F.: The Elements of Statistical Learning, 2nd edn., pp. 337–384. Springer, New York (2009). ISBN 0-387-84857-6
23. Elith, J., Leathwick, J.R., Hastie, T.: A working guide to boosted regression trees. J. Anim. Ecol. **77**, 802–813 (2008)
24. Robert, N., Gary, M., John, E.: Handbook of Statistical Analysis and Data Mining Applications. Academic Press (2009). ISBN-13: 978-0123747655
25. Al-Janabi, S., Salman, M.A., Fanfakh, A.: Recommendation system to improve time management for people in education environments. J. Eng. Appl. Sci. **13**, 10182–10193 (2018).https://doi.org/10.3923/jeasci.2018.10182.10193