

Optimal customer relationship management in telecalling industry by using data mining and business intelligence

T. Kamalakannan^{1*}, P. Mayilvaghanan²

¹ Ph.D, Research Scholar, Department of Computer Science, Vels University, Chennai

² Professor, Department of MCA, Vels University, Chennai

*Corresponding author E-mail: kannan722003@yahoo.co.in

Abstract

Decision making system in telecommunication industries plays a more important role where it is required to find customer churn. Customer churn prediction requires finding out and analyzing the information about the business data intelligence techniques which can be done efficiently by adapting the business intelligence techniques. Business intelligence provides tools to predict and analyze the historical, current and predictive views of business operations. However, this would be more complex task with high volume of data which are gathered from million of telephone users for the time being. It can be handled effectively by introducing the data mining techniques which select the most useful information from the gathered data set from which decision making can be done efficiently. In this research method, telecommunication industry is considered in which customer churn prediction application is focused. The main goal of this research method is to introduce the data mining technique which can select the most useful information from the telecommunication industry dataset. This is done by introducing the Hybrid Genetic Algorithm with Particle Swarm Optimization (HGAPSO) method which can select the most useful information. In this research, the hybrid HGAPSO combines the advantages of PSO and GA optimally. From the selected information, decision making about the customer churn prediction can be done accurately. Finally decision making is done by predicting the customer behaviour using Support Vector Machine classification approach. The performance metrics are considered such as precision, recall, f-measure, accuracy, True Positive Rate (TPR), False Positive Rate (FPR), time complexity and ROC. Experimental results demonstrated that HGAPSO provides highly scalable which is used for prediction examination in the business intelligence.

Keywords: Customer relationship management, Business intelligence, Decision making system, useful information, attribute selection, customer churn prediction.

1. Introduction

While considering the current telecommunication environments, customers have the every right to choose their service providers [1]. At the moment, this is the current status for long-distance, Internet Service Providers (ISP) and wireless service providers. At present competitions are existing in international level in countries like the United Kingdom, Australia and New Zealand. On the other hand, deregulation of the communications market is happening in the U.S. local telephone division, and privatization/deregulation is taking place throughout Europe, Latin America and Asia. Based on the competitive circumstances, the customer is the dominant focus of the carrier's activities [2]. Customer necessities not only regulate service contributions however also transform the network and influence the organizational construction of the carrier to concentrate on specific categories of customer.

In recent times, telecommunication vendors are briskly getting substantial product improvement capabilities because technology transforms determine consumer requirement [3]. On the other hand, they remain to lag behind in realizing the customer. This has paved way for considerable churn as products are developed and rejected in an effort to maintain existing customers and motivate new business. Deregulation and increasing rivalry among other companies is forcing them to move from outdated product-based

operations to consumer-based operations[4]. Customer looks out for new services and lower price services are pushing telecommunications service providers to considerably increasing their efficiency as never before [5].

It is essential that telecommunications vendors must investigate their customers' requirements and modify all their business procedures in the value chain to successfully meet their customers' unique constraints and increasing demands. Based on this argument is the guess that telecommunication organizations have the capability to turn huge volumes of data relating to their customers and services into actionable evidence. With the assistance of business intelligence systems, the companies solve almost all aspects of the value chain to accomplish this objective.

A typical telecommunications company has a huge customer base and varied product offerings. Several telecommunication companies will also service a number of markets across geographies ranging from limited firms serving a combination of urban and big rural communities to huge international vendors serving customers of diverse nationalities and lifestyles [6]. In order to efficiently collaborate with customers and design appropriate offerings, the vendor's CRM approach has to completely utilize the potential of business intelligence solutions. Telco's must influence the massive quantity of data residing in existing systems at each step of the CRM process [7]. The insight gained must subsequently be applied to developing new products

and services to encounter the ever-changing requirements of existing customers and to draw new customers attention. The ever-shrinking boundaries caused through increased competition indicate that telecommunication service providers will have to search for innovative techniques to diminish the cost of implementing effective BI solutions.

The CRM process in a telecommunications company includes three phases:

1. Recognize the most profitable or potentially profitable customers for forthcoming interaction.
2. Realize their requirements and purchasing patterns, and
3. Work together with them in order to meet all of their expectations

In this research method, this is achieved by introducing the HGAPSO algorithm which selects the most useful information from the set of attributes. Thus the optimal decision making about the customer churn can be achieved. The overall research method is implemented and evaluated in the java simulation environment from which it is proved that the proposed technique can provide optimal outcome in terms of accurate information identification results.

The overall organization of the proposed research method is given as following: In this section, detailed discussion about the CRM, and role of BI in CRM is given. In section 2, various related research methodologies has been discussed in detailed. In section 3, proposed research a technique has been discussed in detailed along with suitable examples and explanation. In section 4, experimental evaluation of proposed and existing research method is given. Finally in section 5, conclusion of the research methodology is given based on simulation outcome obtained.

2. Related works

In the present scenario, meeting customer fulfillment is a necessity [8]. The organizations are becoming increasingly aware regarding the benefits of data and information kept in their organization, the requirement to incorporate these HUGE volumes of data and to utilize this information to support the quality of their decision-making, so as to stay at a competitive advantage and to growth profit. Here, a complete literature review for supply chain improvements is discussed and applied to a case study in telecom industry [9]. The literature has mainly concentrated on coordinating pricing and replenishment decisions. This paper offers a proven data analytical scheme for the purpose of identifying variety kinds of characteristics in telecommunication industry.

It is to be observed that, the increased globalization, stronger competition and demand based customers have forced companies to mainly concentrate their attention on the strategic issues of planning, controlling and designing a supply chain as a complete more willingly than on effectiveness and efficiency of separate business functions within its own [10]. Based on true business intelligence, the supply chain manager and further teams and individuals involved in features of the supply chain can generate reports and personalized dashboards and alerts to create objective goals and Key Performance Indicators (KPIs) and monitor shipment systems and further enterprise applications by means of an integrated, single view of data.

While considering [11], it observes that data mining tools are able to discovering patterns in data in limited hours those expert human quantitative analysts might not discover in years of work to assist in making a decision in complex Supply Chain Management (SCM), Customer Relationship Management (CRM) gather and investigate transaction records uninterruptedly. It is mainly concentrated on automated methods for the purpose of extracting patterns or models from data. Telecommunication companies today are working in extremely competitive and challenging atmosphere. Data Mining and BI applications [12] take part a substantial role in the telecommunication business because of the

availability of huge volume of data and the demanding competition in the sector.

In case of [13], it discusses about the main application ranges together with marketing and CRM, Fraud detection and network management. The recent developments in the data mining and BI fields and the implementation and improvement of existing schemes ensure the incessant development and compatibility of telecommunication companies that utilize them. In case of [14], it found that the long-term success of a company depends on the consistency of its suppliers and level of satisfaction of its customers. Collaborative association among customer and supplier has positive significant effect to SCM performance enhancement.

Inventory reduction is one of the foremost goals of SCM [15]. Correspondingly, it is the most frequently shared data among the supply chain partners [16]. As a result, quite a few researchers have discovered the ways to diminish the inventory in a supply chain. On the subject of order fulfillment in [17], it is observed that it was the second most significant supply chain issue in Indian companies. Organizations were remunerating maximum time and consideration to expand order fulfillment. It must be observed that [18] investigates the influence of providing a return policy for unsold goods to two competing retailers facing undefined demand. Adopting the common newsvendor problem model structure and by means of numerical study schemes, the study finds that the provision of a returns policy is reliant on the market conditions faced by the retailers.

The foremost purpose of this review is to explore some foremost issues in business intelligence and SCM literature and to present a framework for classification and analysis. The objective is to give importance to how different subject literatures have contributed work in BI and SCM from different perspectives.

3. Optimal customer relationship management in telecalling industry

In general, the interrelationships with suppliers, distributors, customers and further business that are required to design, construct and retail a product constitute the network of business entities, associations, and processes that is called a supply chain. In specific, Business Intelligence (BI) [19] includes the incorporation of core information with significant business information for the purpose of detecting significant events, determine new business circumstances and expect business situations. It comprises the capability to observe business trends, to change and adjust quickly as situations transform and to make intelligent business decisions on uncertain findings and inconsistent information. BIschemes are finding their utilization in several fields like: government, retail, financial services and telecommunication industry. In the midst of these, telecommunications is chosen to demonstrate the role of BI in this particular field. BI consequently does effectively increase profitability in the telecommunication industry and can assist management to define strategies, search for opportunities, support decision making, recognize complications and substantiate actions.

The main goal of this research method is to introduce the data mining technique which can select the most useful information from the telecommunication industry dataset. This is done by introducing the Hybrid Genetic Algorithm with Particle Swarm Optimization (HGAPSO) method which can select the most useful information. In this research, the hybrid HGAPSO combines the advantages of PSO and GA optimally. From the selected information, decision making about the customer churn prediction can be done accurately.

3.1. System overview

The overall performance of proposed system is illustrated in fig1. It is shows that the initial phase is data preparation, it is collected the data, integrated and cleaned. The data preprocessing is done

using Normalization based K means algorithm. Then the attributes are selected using HGAPSO algorithm to search for relevant attributes, eliminate irrelevant or redundant ones.

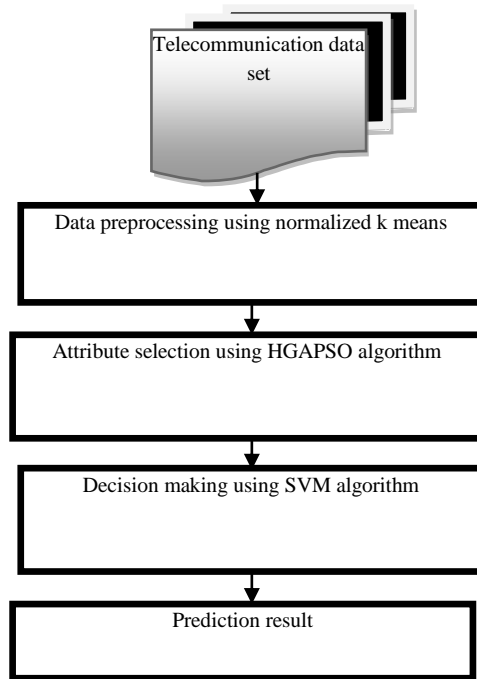


Fig. 1: Overall architecture of proposed system

3.2. Data preprocessing using normalized K means algorithm

Clustering is defined as grouping set of points into clusters according to distance measure. The result of clustering is set of clusters each cluster will have set of points with small distance from one another and with large distance from other clusters. This technique can detect outliers, since it grouping similar points into a cluster while points that fall out the clusters are considered as outlier points. Data Mining can produce efficient results if normalization is effectively applied on the dataset. It is a process utilized to regulate all the attributes of the dataset and provide them equal weight in order that unwanted or noisy objects can be removed and there is valid and consistent data which improves the accuracy of the result. K-Means scheme utilizes Euclidean distance that is extremely disposed to irregularities in the size of numerous characteristics. There are a number of data normalization schemes like Min-Max, Z-Score and Decimal Scaling. The best normalization scheme is completely based on the data to be normalized. At this time, Min-Max normalization scheme is used in this algorithm, since the taken dataset is limited and has not considerable inconsistency between minimum and maximum. Min-Max normalization scheme executes a linear transformation on the data. In this scheme, the data is fitted in a predetermined boundary or in a predetermined interval.

A uniform scheme is employed for the purpose of finding score through taking the average of the attribute of each data point which will produce initial centroids that follow the data distribution of the particular set. A sorting scheme is implemented to the score of each data point and subsequently divided into k subsets, in which k represents the number of clusters. To end with, the nearest value of mean from each subset is taken as preliminary centroid. In this scheme, a weight is introduced with each attribute, which makes the scheme beneficial as it can cause improvement of any feature of the dataset by means of increasing the weight associated with that attribute. The algorithm is given in Fig 1:

Algorithm 1: Steps of N-K means Algorithm

INPUT: A dataset with d dimensions

OUTPUT: Clusters

1. Provide initial data set.
2. Discover the maximum and minimum values of each feature from the dataset.
3. Normalize real scalar values of datasets with extreme values by means of equation:

$$v' = (v - \min(e)) / (\max(e) - \min(e)) \quad (1)$$
 In which, $\min(e)$ and $\max(e)$ represents the minimum and the maximum values for attribute E .
4. Pass the number of clusters and generate initial centroids using algorithm 2.
5. Produce clusters

Algorithm 2: Initialization of centroids

1. Compute the average score of each data point.
 - 1) $d_i = x_1, x_2, x_3, x_4 \dots x_n$
 - 2) $d_i(\text{avg}) = (w_1 * x_1 + w_2 * x_2 + w_3 * x_3 + \dots + w_m * x_m) / m$, in which x represents attribute's value, m denotes number of attributes, w indicates weight to multiply to confirm fair distribution of cluster.
2. Organize the data in accordance with the average score.
3. Segment the data by means of k subsets.
4. Compute the mean value of each subset.
5. Take the nearest possible data point of the mean as the initial centroid for each data subsets.

3.3. Attribute selection using HGAPSO

For prediction of specific results attribute selection takes part a significant role in increasing the prediction and decreases the curse dimensionality problem. Attribute selection aims to discover the most vital information from a particular set of attributes. As this task can be seen as an optimization complication, the combinatorial development of the possible solutions possibly be in-viable for an exhaustive search. In this research, 20 categories of attributes are considered for instance, Gender, Senior Citizen, Partner, Dependent, tenure, Phone Service, Multiple Lines, Internet Service, Online Security, Online Backup, Device Protection, Tech Support, Streaming TV, Streaming Movies, Contract, Paperless Billing, Payment Method, Monthly Charges, Total Charges, Churn. These attributes are contained information about Customers who left within the last month—the column is called Churn. Services that each customer has signed up for—phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies. Customer account information—how long they've been a customer, contract, payment method, paperless billing, monthly charges, and total charges. The demographic information of customers such as—gender, age range, and also checks if they have partners and dependents. These attributes are categorized as numerical and categorical. The categorical attributes are contained 'yes' or 'no' and '0' or '1' etc., so important attributes or variables of the customer are selected based on the HGAPSO algorithm.

This new evolutionary learning scheme completely depends on a hybrid of Genetic Algorithm (GA) and Particle Swarm Optimization (PSO), and is thus called HGAPSO. In HGAPSO, individuals in a new generation are created, not only through crossover and mutation operation as in GA, however also by PSO. The conception of elite strategy is adopted in HGAPSO, in which the upper-half of the best-performing individuals in a population are regarded as elites. On the other hand, instead of being reproduced directly to the subsequent generation, these elites are enhanced initially. The group constituted by the elites is regarded as a swarm, and each elite corresponds to a particle within it. Based on this, the elites are enhanced through PSO, an operation which mimics the maturing phenomenon in nature. These enhanced elites constitute half of the population in the new

generation, whereas the other half is generated by performing crossover and mutation operation on these enhanced elites.

In the first stage of solving the problem of optimization the PSO scheme will generate an initial population near the global optima. Following that the algorithm switches to GA and GA takes this initial population and continues to solve the optimization problem. The step by step algorithm for the proposed attribute selection using HGAPSO is given below:

Step 1. The number of attributes declared and the decision making is carried out.

Step 2. The initial population of individuals is generated satisfying the CRM constraints and also it is verified that only one attribute is taken in each line.

Step 3. For each individual in the population, the fitness function given by (7) is assessed after running attributes.

Step 4. The velocity and new population is updated by (11).

Step 5. If maximum iteration number is reached, then go to next step else go to step 3

Step 6. Get the last population as initial population and using GA update the population

Step 7. For each individual in the population, the fitness function given by (8) is evaluated after running attributes.

Step 8. If the stop criterion is satisfied, then go to step 9 else go to step 6.

Step 9. Output the results

3.4. Decision making using SVM

The customer churn separation or prediction is examined using Support vector machine. Support vector machine (SVM) is a supervised algorithm and it is used for prediction purpose in any given dataset. For customer churn prediction, we use SVM which is optimal separating hyperplane between the two classes of data. SVM models used to produce better prediction results.

Formula

Training dataset (D)

$$D = \{x_i, y_i\}_{i=1}^N, x \in R^n, y \in \{-1, 1\} \quad (1)$$

D is training dataset, x and y is input variables

$$y^i [|w^T x^i + b|] \geq 1, i=1 \text{ to } N \quad (2)$$

w^T and b are separated variables

To reduce the error minimization we can use given below formula

$$\Phi(w) = \frac{1}{2} \|w\|^2 \quad (3)$$

Estimating function

$$F(x) = \sum_{i=1}^{N_{sv}} (x_i, y_i) k(x_i, y_i) + b \quad (4)$$

SVM algorithm procedure

Given telecommunication dataset $D=(x_1, y_1), \dots, (x_n, y_n)$, $C// x$ and y –labeled samples and C -class

Initialize vector $v=0$, $b=0$; class// v -vector and b -bias

Train an initial SVM and learn the model

For each $x_i \in X$ do// x_i is a vector containing features describing example i

Classify x_i using $f(x_i)$

If $y_i f(x_i) < 1$ // prediction class label

Find w', b' for known data// w', b' for new features

Add x_i to known data

Minimize the error function using (3) and estimate using (4)

If the prediction is wrong then retrain

Repeat

End

Classify attributes as normal or abnormal

4. Experimental results

In this section, in order to evaluate the performance of HGAPSO attribute selection used a training set of customer information

collected over a six-month time period. Each customer is categorized into one of two predetermined groups and his/her churn propensity is observed and revised according to his/her most recent three-month information. In this way, it is uncomplicated to simulate the real-world setting of churn prediction. A churn prediction system is supposed to be measured through its capability for the purpose of recognizing churners for marketing use, and in this work effectively employed the Receiver Operating Characteristic (ROC) curve and top-quantile-lift values to provide a complete evaluation of HGAPSO scheme. Subsequently, the results with a GA, and PSO without customer separation is compared.

4.1. Data collection

In this research work, a data set is selected from a telecommunication organization which comprises a segment of mobile customers. The dataset is IBM Watson Analytics Telco Customer Churn data from <https://community.watsonanalytics.com/predictive-insights-in-the-telco-customer-churn-data-set/>. This data set provides information of behavior to retain customers. A telecommunications company is concerned about the number of customers leaving their landline business for cable competitors. They need to understand who is leaving. And they have to analyst at this company and find out who is leaving and why. The data set includes information about: Customers who left within the last month—the column is called Churn. Services that each customer has signed up for—phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies. Customer account information—how long they've been a customer, contract, payment method, paperless billing, monthly charges, and total charges.

4.2. Evaluation criteria

In this study, the Area Under receiver Curve (AUC), sensitivity, and specificity are used to quantify the accuracy of the predictive models. If True Positives (TP), False Positives (FP), True Negatives (TN) and False Negatives (FN) are the TP, FP, TN and FN in the confusion matrix, then the sensitivity is $(TP/(TP + FN))$: The proportion of positive cases which are predicted to be positive.

The specificity is $(TN/((TN + FP)))$: The proportion of negative cases which are predicted to be negative [19]. To assess the accuracy of a classifier independent of any threshold, ROC analysis can be used. The horizontal axis and the vertical axis of an ROC curve are defined by Equations 5 and 6 respectively [19].

$$x=1 - \text{specificity (t)} \quad (5)$$

$$y=\text{sensitivity (t)} \quad (6)$$

To measure the accuracy of a model, the AUC can be measured [19, 20].

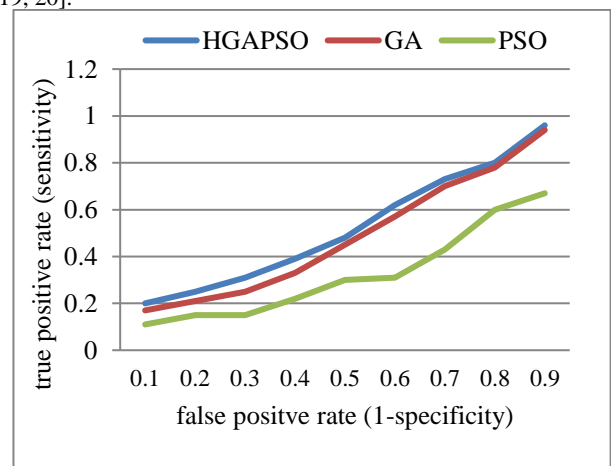


Fig. 2: ROC curves of the separation

Fig. 2 shows the performance of the separation, with the cutoff point for best sensitivity and specificity. This HGAPSO model is then applied back to the training data and generates customer samples. The attribute selection of proposed HGAPSO schema is 96 %, for specificity value of 90 % training sample.

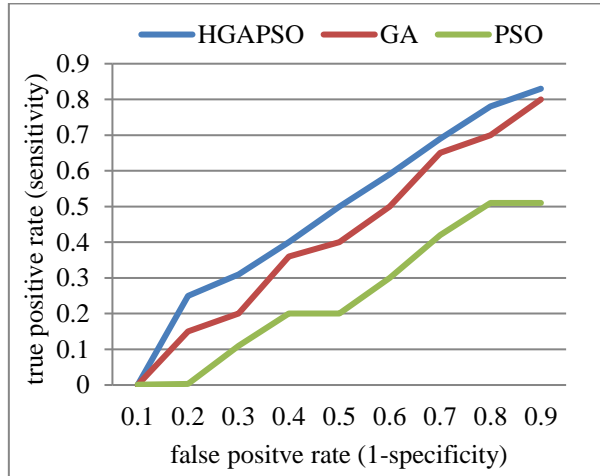


Fig. 3: ROC curves of predictions on samples

Fig. 3 shows that, for customers who belong to the training samples, the ROC curve of proposed churn prediction model is located above models, with the Curve increased from 80.13 % to 83%. HGAPSO model achieves a much better result than existing GA and PSO methods.

Processing time comparison

The HGAPSO prediction model discovers frequent churn prediction and much greater efficiency than the existing prediction model of GA and PSO shown in Fig.4. The proposed HGAPSO prediction model takes less computation time to predict the customer churn when compared to existing system. The HGAPSO is reportedly working efficiently and in many cases, it's much faster than GA and PSO.

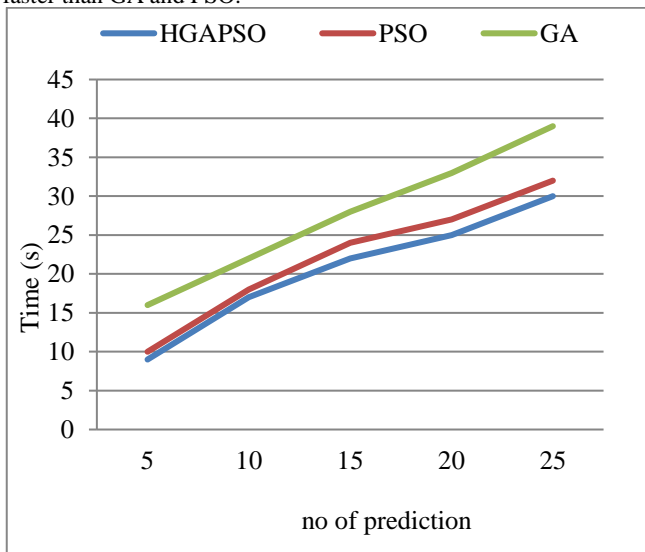


Fig. 4: Processing time comparison

Accuracy comparison

The HGAPSO prediction model discovers frequent churn prediction and much greater accuracy results than existing prediction model of n Fig 5. When the number of prediction is increases the accuracy of the result is increases. The proposed HGAPSO produces high accuracy rate when compared to existing system. With the Area Under receiver Curve increased from 98.13

% to 98.3%. The HGAPSO is reportedly working efficiently and in many cases produces high accuracy rate than GA and PSO.

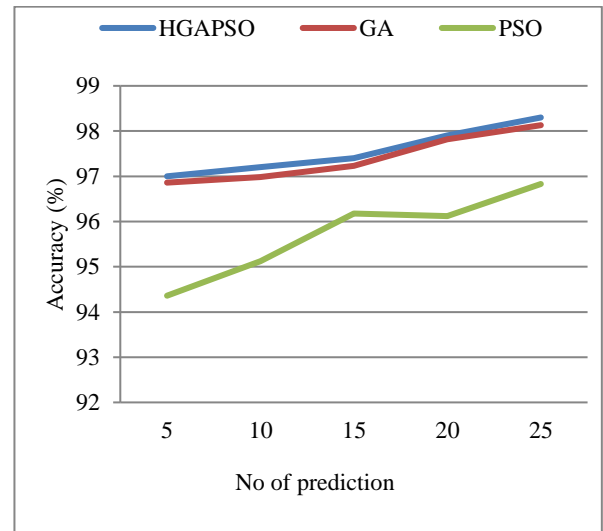


Fig. 5: Accuracy comparison

5. Conclusion

In this research method, telecommunication industry is considered in which customer churn prediction application is focused. The main goal of this research method is to introduce the data mining technique which can select the most useful information from the telecommunication industry dataset. This is done by introducing the Hybrid Genetic Algorithm with Particle Swarm Optimization (HGAPSO) method which can select the most useful information. In this research, the hybrid HGAPSO combines the advantages of PSO and GA optimally. From the selected information, decision making about the customer churn prediction can be done accurately. The performance metrics are considered such as precision, recall, f-measure, accuracy, True Positive Rate (TPR), False Positive Rate (FPR), time complexity and ROC. Experimental results demonstrated that HGAPSO provides highly scalable which is used for prediction examination in the business intelligence.

References

- [1] Rashidirad M, Salimian H, Soltani E & Fazeli Z, "Competitive strategy, dynamic capability, and value creation: Some empirical evidence from UK telecommunications firms", *Strategic Change*, Vol.26, No.4, (2017), pp.333-342.
- [2] Zhao ZY, Tang C, Zhang X & Skitmor M, "Agglomeration and competitive position of contractors in the international construction sector", *Journal of Construction Engineering and Management*, Vol.143, No.6, (2017).
- [3] Purdy C, Evelyne, ROCH, Schubert D & Strong P, "U.S. Patent Application No. 15/046,039", (2016).
- [4] Pucciarelli F & Kaplan A, "Competition and strategy in higher education: Managing complexity and uncertainty", *Business Horizons*, Vol.59, No.3, (2016), pp.311-320.
- [5] Brockhoff K, *Customer Integration into Continuous Development of IT-based Services*, The Palgrave Handbook of Managing Continuous Business Transformation, (2017), pp.315-334.
- [6] Walo MT, "Net chain: An analytical approach to local economic development", *The Marketing Review*, Vol.16, No.1, (2016), pp.46-61.
- [7] Wynn M, Turner P, Banik A & Duckworth G, "The impact of customer relationship management systems in small business enterprises", *Strategic Change*, Vol.25, No.6, (2016), pp.659-674.
- [8] Pizam A, Pizam A, Shapoval V, Shapoval V, Ellis T & Ellis T, "Customer satisfaction and its measurement in hospitality enterprises: a revisit and update", *International Journal of Contemporary Hospitality Management*, Vol.28, No.1, (2016), pp.2-35.

- [9] Yawar SA & Seuring S, "Management of social issues in supply chains: a literature review exploring social issues, actions and performance outcomes", *Journal of Business Ethics*, Vol.141, No.3, (2017), pp.621-643.
- [10] Helo P, Gunasekaran A & Rymaszewska A, "Improving Marketing and Operations Strategy through Industrial Services", *Designing and Managing Industrial Product-Service Systems*, (2017), pp.37-42.
- [11] Singh H, "Implementation Benefit to Business Intelligence using Data Mining Techniques, Computer Faculty, Education Department Punjab", *International Journal of Computing & Business Research*, (2012).
- [12] Vercellis C, *Business Intelligence Data Mining and Optimization for Decision Making*, Wiley, (2009).
- [13] Joseph V, "Data mining and business intelligence applications in telecommunication industry", *International Journal of Engineering and Advanced Technology*, Vol.2, No.3, (2013).
- [14] Daspit JJ, Holt DT, Chrisman JJ & Long RG, "Examining family firm succession from a social exchange perspective: a multiphase, multistakeholder review", *Family Business Review*, Vol.29, No.1, (2016), pp.44-64.
- [15] Pagelet. al., "Supply chain management in electronics & telecommunication industry-a literature survey", (1999).
- [16] Kariuki SW & Rotich G, "Role of Stores Management in Reduction of Redundant Stock, A Case Study Of Keroche Breweries Limited, Kenya", *International Journal of Project Management*, Vol.1, No.5, (2017), pp.80-97.
- [17] Jajja MSS, Jajja MSS, Kannan VR, Kannan VR, Brah SA, Brah S A & Hassan SZ, "Supply chain strategy and the role of suppliers: evidence from the Indian sub-continent", *Benchmarking: An International Journal*, Vol.23, No.7, (2016), pp.1658-1676.
- [18] Yao Z, Leung SC & Lai KK, "Analysis of the impact of price-sensitivity factors on the returns policy in coordinating supply chain", *European Journal of Operational Research*, Vol.187, No.1, (2008), pp.275-282.
- [19] Castellanos M & Dayal U, "Business Intelligence for the Real-Time Enterprise", *Springer*, (2008).