

Research Article

Fuzzy particle swarm optimization (FPSO) based feature selection and hybrid kernel distance based possibilistic fuzzy local information C-means (HKD-PFLICM) clustering for churn prediction in telecom industry



C. K. Praseeda¹ · B. L. Shiyakumar²

Received: 15 July 2020 / Accepted: 15 April 2021

Published online: 10 May 2021 © The Author(s) 2021 OPEN

Abstract

Customer churn has been considered as one of the key issues in the operations of the corporate business sector, as it influences the turnover directly. In particular, the telecom industries are seeking to develop new approaches to predict potential customer to churn. So, it needs the appropriate algorithms to overcome the increasing problem of churn. This work proposed a churn prediction model that employs both strategies of classification and clustering, that helps in recognizing the churn consumers and giving the reasons after the churning of subscribers in the industry of telecom. The process of information gain and fuzzy particle swarm optimization (FPSO) has been executed by the method of feature selection, besides the divergence kernel-based support vector machine (DKSVM) classifier is employed in categorizing churn customers in the proposed approach. In this way, the compelling guidelines on retention have generated since the process plays a vital role in customer relationship management (CRM) to suppress the churners. After the classification process, the churn customers are divided into clusters through the process of fragmenting the data of churning customer. The cluster-based retention offers have provided by the clustering algorithm of hybrid kernel distance-based possibilistic fuzzy local information C-means (HKD-PFLICM), whereas the measurement of distance have accomplished through the kernel functions such as the hyperbolic tangent kernel and Gaussian kernel. The results reveal that proposed churn prediction model (FPSO- DKSVM) produced better churn classification results compared to other existing algorithms such as K-means, flexible K-Medoids, fuzzy local information C-means (FLICM), possibilistic FLICM (PFLICM) and entropy weighting FLICM (EWFLICM).

Article highlights

- Customer churn is a major concern in most of the companies as it influences the turnover directly.
- The performance of churn prediction has been improved by applying artificial intelligence and machine learning techniques.

 Churn prediction plays a crucial role in telecom industry, as they are in the position to maintain their precious customers and organize their Customer Relationship Management.

Keywords Customer relationship management (CRM) \cdot Churn prediction \cdot Retention \cdot Telecom \cdot Clustering \cdot Classification \cdot Feature selection

☑ C. K. Praseeda, prasee2000@yahoo.com | ¹Bharathiar University, Coimbatore, India. ²Sri Ramakrishna College of Arts and Science, Coimbatore, India.



SN Applied Sciences

(2021) 3:613

| https://doi.org/10.1007/s42452-021-04576-7

1 Introduction

In developed countries, the telecom industry has been established as one of the leading industrial sectors, in which the level of competitions is in peak, in terms of technology advancements as well as the growing number of service providers [1]. Industries contribute the great effort based on various approaches to ensure their endurance in this competitive field, among which three primary approaches are presented to increase the turnover namely, (i) acquire new customers, (ii) upsell existing customers, and (iii) increase customer's retention period. Nevertheless, it can be concluded that the third strategy is foremost beneficial, while assessing all the three approaches by considering each of their Return on Investment (Rol) value [2]. Moroever, it shows that the existing customer retention cause lesser expense than obtaining new customer [3]. Additionally, it requires solely a minimal effort than upselling approach [4]. Industries are required to diminish the possibility of customer churn in terms of enforcing this approach, namely "the customer movement from one provider to another" [5]. Churn prediction plays a crucial role in telecom industry, as they are in the position to maintain their precious customers and organize their Customer Relationship Management (CRM) [6, 7]. CRM is a wide-ranging system that regulates the process of structuring, administrating, and reinforcing the reliable and enduring relationships with customers, which is widely accepted strategy and implemented in various sectors such as, retail market, insurance, banking, telecommunication, etc. This strategy primarily intended to retain the customers and the necessity of the intention is apparent that the expenditure in obtaining new customers is extensively higher when compared to the expense of customer retention (multiplied the expense 20 times in some instance). Hence, efficient tools are necessary to develop and enforce the customer retention systems (churn models) and to deal with applications for Business Intelligence (BI). The churning can be caused by the competitive marketing factors like, insufficient service to fulfil customer expectations, competitive strategies, attractive products, new regulations, etc. Churn strategies intended to spot the initial indications of churn and identify the customers with the probability of leaving. In order to classify those customers who are susceptible to churn, it necessitates the system of churn prediction to be created. To procure the optimal forecast on customer churn, several strategies of data mining, like Neural Networks (NNs), Clustering, Decision Tree (DT) [8], Regression [9], Support Vector Machine (SVM) [10], and the ensemble of Hybrid methods [11] were employed in existing studies. Industries have realized further vigorous tactics with the help of technical advancements, in order to guarantee the maximum level of churn customers remains in their company. The focus of the researchers is on discriminating the customers to find out those who are supposed to be churn to other operators [12]. There are heavy competitions in the sector, by reason of deregulation in telecom industry and consumers have lot of options to choose. Hence, the scenario demands the telecom operators to absorb the customer requirements and amend it, which has been considered as a primary objective for them to keep their customers away from the competitor [13]. The key intention of the present studies is that the telecom data needs to be utilized in huge quantity to classify the precious churn consumers. Nonetheless, the current approaches are bound with a lot of constraints and challenges in a real-world environment. The process apparently includes the issue of data derivation with missing values, as it generates huge quantity data in the telecom industry, which in turn the least efficient results from prediction systems. Therefore, the approaches of data preprocessing is developed to resolve these problems, which eradicates the data noise and enable the approach to categorize the data in a proper manner, and eventually, with better efficiency.

Even though the study has employed the feature selection approach, numerous instructive features are ignored during the development process of the system model [14]. Statistic-based approaches have been largely applied in different fields, which source the inefficient outcomes from predictive strategy. The current research verifies their approaches through standard datasets [15, 16]. But, it lacks to provide the appropriate depiction of data and that in turn does not make sense in decision making. In churn prediction approach, the feature selection method aids to choose the essential features like, information gain, and correlation attribute ranking filter [17]. Two massive datasets from the telecom domain are processed through churn and non-churn classification by applying several machine learning strategies [17]. In this research work, Information Gain (IG) and Fuzzy Particle Swarm Optimization (FPSO) techniques are used in this paper, concerning the enhancement of feature selection efficiency. Churn customer data has been categorized through applying the algorithm of Divergence Kernel-based Support Vector Machine (DKSVM). Besides, by utilizing Hybrid Kernel Distance-based Possibilistic Fuzzy Local Information C-Means (HKD-PFLICM) method, the customers are divided into three clusters such as Low, Medium, and Risky on the basis of customer activities, through which the customer profiling has been executed. The evaluation metrics such as F-measure, True Positive Rate (TPR), Precision, False Positive Rate (FPR), Recall and Receiver Operating Characteristic (ROC) have been utilized to evaluate the performance of proposed churn prediction approach. The rest of the paper is arranged as follows, Sect. 2 discuss few related methods proposed earlier. Section 3 illustrates proposed methodology in detail. Section 4 provides information about the performance evaluation of proposed approach in comparison with few other methods discussed in Sect. 2. Section 5 concludes the papers with insight for future research direction.

(2021) 3:613

2 Literature review

A number of techniques such as machine learning, data mining, hybrid techniques etc. are being utilized by the researchers in Churn prediction. As a result of the research, many models confining to customer churn prediction are innovated, this in turn aids in taking decision and CRM for identification, prediction, and retainment of churning customers by the companies.

Huang and Kechadi [14] established a customer behaviour prediction by incorporation of supervised and unsupervised techniques via novel hybrid model-based learning system. This technique comprises of modified k-means clustering algorithm and classic First-Order Inductive Learner (FOIL) technique which takes into account various factors such as customer's behaviour, usage pattern and network operations. The datamining pre-processing is the challenging concern for detection of outliers and redundant data examples. Hence, this hybrid model helps in achieving better clustering result. Tsai and Lu [18] integrated back-propagation Artificial Neural Networks (ANNs) and Self-Organizing Maps (SOM) techniques for creation of hybrid model and it offers good prediction accuracy results comparatively in line with single NN baseline model and Types I and II errors over three kinds of testing sets.

Jahromi et al. [19] utilized clustering phase and classification phase which is a dual-step model building methodology. The RFM related features have a key role in classification of the customer into four clusters thereby churn logical definition is obtained; additionally, algorithms play a significant role in churn prediction model extraction in clusters formed. The gain measure evaluation supports in differentiating the performance of employed algorithms which also suggest that multi-algorithm methodology offers more advantages compared to single-algorithm. This research is prolonged to prepaid mobile phone organizations [19] with the intention of determining responses to customer loyalty results. Verbeke et al. [20] utilized a methodology involving data mining in customer churn prediction modeling. Comprehensibility and intuitiveness of churn prediction models are the most important factors in decision. The traditional rule induction classifiers C4.5 and Repeated Incremental Pruning are overcome by the novel data mining techniques which are pragmatic to churn prediction modelling for establishing an Error Reduction (RIPPER). The novel techniques namely Ant-Miner + and Active Learning Based Approach (ALBA) offer good accuracy comparatively. The Ant Colony Optimization (ACO) based AntMiner + is one of the most promising data mining technique and has option of encompassing domain knowledge by stating monotonicity constraints on final rule-set. ALBA exhibits high predictive accuracy of non-linear support vector machine model with unambiguousness of rule-set format. Lu et al. [2] proposed a strategy for enhancement of customer churn prediction model namely boosting which allows customer categorization on the basis of algorithm weight assignment. This helps in determining the higher risk customer cluster. The basis learner in this research is Logistic Regression (LR) and a churn prediction model construction is accomplished on every cluster respectively whose results are contrasted with a single LR model. It also exhibits good separation of churn data which is more useful for churn prediction analysis. Zhang et al. [21] suggested a prediction methodology on the basis of interpersonal influence which incorporates propagation process and customers' personalized characters. Due to this integration, the novel techniques exhibit good prediction accuracy in contrast with traditional method. Data mining algorithm is a key concern in customer history analysis and prediction. Qureshi et al. [4] projected a data mining technique for churn customer identification. The patterns are obtained for determining probable churners depending on the historical data. The significant algorithms used in this work are Regression analysis, Decision Trees (DTs) and ANNs.

Mishra and Reddy [22] utilized Convolutional Neural Network (CNN) for accomplishing churn prediction which exhibits better accuracy. The experimental results provide accuracy of 86.85%, precision 91.08, error rate of 13.15%, recall 93.08%, F-score 92.06%. The tensor flow package with respect to time and accuracy can be still utilized for enhancement of the model performance. Amin et al. [23] extracted significant decision rules associated to customer churn and non-churn by utilizing a Rough Set Theory (RST) based decision-making technique. The process is carried out by classifying the churn and non-churn customers in addition to customers who conceivably may churn in the proximate prospect. RST based on Genetic Algorithm (GA) [24] is considered as greatest efficient system for extracting implicit knowledge in form of decision rules from publicly available, benchmark telecom dataset.

Mitrovi¢ et al. [25] involve daPareto multi-criteria optimization for the determination of optimal feature type combinations which is a reusable scheme which helps in sorting the gap amid predictive performance and operational efficiency. Caigny et al. [26] presented a new hybrid algorithm namely Logit Leaf Model (LLM) for improved classification data. There are two stages specifically segmentation phase and prediction phase in LLM. Decision is a key factor in decision rules identification in case of first stage and a model creation is done for every leaf of this tree. This is considered as the proficient hybrid technique due to predictive performance and comprehensibility in contrast with DTs, LR, RF and LR trees. Few key benefits are also attained due to the comprehensibility feature in contrast with decision trees or logistic regression. Stripling et al. [27] introduced ProfLogit for maximizing the Expected Maximum Profit measure for customer Churn (EMPC) in training step using a GAs. Also, there occurs a similarity of interior model structure with respect to lassoregularized model. The EMPC basis helps in obtaining the threshold-independent recall and precision measures depending on expected profit maximizing fraction.

SN Applied Sciences

Agrawal et al. [28] utilized Deep Learning Approach for churn on a Telco dataset prediction. A non-linear classification model is constructed by means of a Multilayered Neural Network. This model takes into account various features such as customer, support, usage and contextual features. The determination parameters along with the churn probability are estimated. The final weights on these features are greatly utilized by the trained model for prediction of churn. Amin et al. [29] designed a new Cross-Company Churn Prediction (CCCP) model for explicitly for telecommunication sector. Systematically the exploration of CCCP model performance is accomplished with respect to data transformation technique and state-of-the-art classifiers.

A novel fuzzy particle swarm optimization (NFPSO) is explained by Tian & Li [30]. In this work, based on the control information translated from Fuzzy Logic Controller (FLC), learning coefficient and inertia weight are adjusted adaptively during search process. The Canonical Particle Swarm Optimization (CPSO) is introduced with two-input and two-output FLC. Three benchmark functions are used in experimentation and proposed NFPSO's effectiveness is demonstrated with related metrics.

With robust update mechanisms and chaos-based initialization, modified particle swarm optimization is introduced by Tian and Shi [31]. For generating uniformly distributed particles, Logistic map is used and it enhances initial population's quality. Sigmoid-like inertia weight is formulated for making PSO to adapt inertia weight adaptively. For enhancing swarm diversity, particles with fitness value less than average fitness value are applied with wavelet mutation. In addition, global best particle is applied with an auxiliary velocity-position update mechanism exclusively. The MPSO's convergence is guaranteed effectively because of this mechanism. Extensive experimentation on CEC'13/15 test suites and standard image

segmentation validates the MPSO algorithm's efficiency and effectiveness.

A new technique to enhance PSO is implemented by Neshat [32]. For each iteration of proposed FAIPSO technique, for every particle, acceleration coefficients are adjusted adaptively. A fuzzy inference system is used to control the acceleration coefficients adaptively.

Ullah et al. [17] introduced a churn prediction model involving classification and recognition of the churn customers by providing the factors behindhand the telecom sector churning customers. The information gain and correlation attribute ranking filter are key factors in Feature selection. The rules created by expending the attribute-selected classifier algorithm are also equally important factor in churning of customers. The proposed work is motivated from this work by presenting new feature selection and classifier. Swarm based algorithms is introduced instead of filtering based feature selection. By analysis the above existing approaches, in today's competitive world, to keep customers satisfied is a key to success for telecommunication companies.

3 Proposed methodology

The proposed work concentrates on presenting a churn prediction model. Initially, data preprocessing is accomplished for elimination of noise, imbalanced data features along with data normalization. The next step involves the significant feature selection carried out by Information Gain (IG) and Fuzzy Particle Swarm Optimization (FPSO). Thirdly, Divergence Kernel based Support Vector Machine (DKSVM) classifier has been proposed for customer churn prediction. DKSVM classifier is introduced for churn and non-churn classification of telecom sector dataset in addition to which factors are recognized for utilizing in clustering algorithms involved in the subsequent step. The next step involves Hybrid Kernel Distance based Possibilistic Fuzzy Local Information C-Means (HKD-PFLICM) for customer profiling operation. The HKD-PFLICM segregates into three groups Low, Medium and Risky customers based on their behaviour [33]. The concluding step involves the retention strategies for every category of churn customers by means of model recommendations. The churn prediction accuracy is estimated utilizing the metrics such as True Positive Rate (TPR), Precision, False Positive Rate (FPR), Recall, Receiver Operating Characteristic (ROC) area and F-measure. Figure 1 shows theproposed model for customer churn prediction.

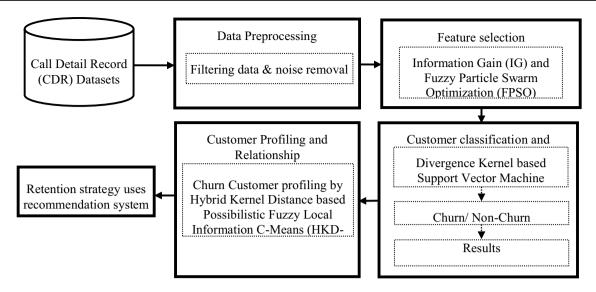


Fig. 1 Proposed model for customer churn prediction

3.1 Call detail record (CDR) datasets

This research utilizes two datasets. The customer churn prediction problem is analysed with the help of first dataset (telecom churn (cell2cell)) acquired from Teradata center for customer relationship management at Duke University. Pre-process operation is accomplished for Cell2Cell dataset and for analyzing Process, a balanced version is delivered comprising 71,047 instances and 58 attributes. The first dataset is a publicly existing telecom churn (cell2cell) "https://www.kaggle.com/jpacse/datas ets-for-churn-telecom/version/2". The second dataset is a publicly accessible churn-bigml dataset "http://github. com/caroljmcdonald/mapr-sparkmlchurn/ tree/master/ data". There are 3333 instances and 21 features in the dataset and the representation of targeted churn customer's class is given by "T" constituting 14.5% of total data while 85.5% constitutes to non-churn customers represented as "F". Table 1 designates about the above datasets.

3.2 Noise removal

The noise in data is more significant since it makes the data useless which in turn affects the results. The telecom data-set comprises lot of missing values, incorrect values like "Unknown" attributes etc. There are 21 features involved in second dataset. The filtering of dataset and features is carried out to make sure of only useful features. The features possessing the unknown range are removed from the dataset. Just remove the unknown samples from the dataset.

3.3 Feature selection—information gain (IG) and fuzzy particle swarm optimization (FPSO)

The relevant features selection from a dataset is accomplished on the basis of domain knowledge which is considered to be a critical step namely Feature selection. Various researchesare being carried out in feature selection involved in churn predictions [34, 35]. Information Gain (IG) and Fuzzy Particle Swarm Optimization (FPSO) for feature selection are greatly utilized for feature selection in this research. The information gain and accuracy of the FPSO algorithm aids in extracting 39 features from the total 58 features in first dataset.

3.4 Information gain (IG)

The main scope of Information Gain (IG) is the machine learning area which pertains to entropy-based feature evaluation method. The Information Gain is defined as the amount of data delivered by the feature for the churn prediction adopted in feature selection. Information gain gives the estimation of term utilized for data classification for measuring the data prominence for the churn prediction and estimates the attribute relevance (A in class C). The higher the relevance amid classes C and attribute A is indicated by higher value of mutual information between classes C and attribute A. The information gain expression is presented further down in Eq. (1),

$$IG(C, A) = H(C) - H(C|A)$$
(1)

$$H(C) = \sum_{c \in C} p(C)logp(C)$$
 (2)

where H(C) is entropy of class revealed further down in Eq. (2) and H(C|A) is conditional entropy of the class given attribute is revealedfurther down in Eq. (3),

$$H(C|A) = -\sum_{C \in C} p(C|A) \log p(C|A)$$
(3)

$$IG(C, A) = 1 - H(C|A)$$
(4)

Attribute A and classes C are not connected in any way, expressed by, if only if H(C|A) = 1, then the least value of I (C, A) occurs. Conversely, incline to select attribute A which probably occurs in one class C whether as positive or negative. That is to say, optimum features are set of attributes which solely occur in one class. Signifies that, when the value of P (A) is similar to the value of P (A \mid C₁), the highest IG(C, A) is reached, consequently both values of $P(C_1 | A)$ and H(C₁| A) being 0.5. Whereas, if values of P (A) and P $(A|C_1)$ are similar, then the value of $P(A|C_2)$ become $P(A|C_2)$ = 0 and $H(C_1|A) = 0$. The IG(C,A) value is differed from 0 to 0.5. The Advantages of Information gain ratio is biasing decision tree against considering attributes with large number of distinct values. So, it solves drawback of information gain—namely, information gain applied to attributes that can take on large number of distinct values might learn training set too well.

 $x_i(t)$ represents the particle position i in the time (t), and $ve_i(t+1)$ expresses the particle velocity i in the time (t+1)for feature selection in churn prediction. The gbest PSO version was used throughout this study, in which each particle's neighborhood experience has received from the overall instances (swarm) for feature selection in churn prediction. The following calculation (6) assess the velocity for each particle i,

$$ve_{ij}(t+1) = ve_{ij}(t) + ac_1 ra_{1j}(t) [y_{ij}(t) - x_{ij}(t)]$$

$$+ ac_2 ra_{2j}(t) [\hat{y}_{ij}(t) - x_{ij}(t)]$$
(6)

where ve_{ii}(t) expresses the particle velocity i in dimension j in time t, $x_{ii}(t)$ signifies the particle's location in time t, y_{ii}(t) represents the particle's optimal position (feature position), $\hat{y}_{ii}(t)$ depicts the overall optimal position (feature position) of entire particles, ac₁, ac₂ are the acceleration constant for the best chosen local and global features in the churn prediction, and in the interval ra_{1i}, ra_{2i} are random numbers [0, 1].

In recent times, some researches intended to develop the strategies in order to augment the convergence and effective solutions have derived with the standard PSO, such as the velocity improvement, inertia weight. The inclusion of inertia weight improves the process of modifying best PSO version (ac_2). Apart from managing the particle's velocity and direction, the inertia weight enable the control over exploration and exploitation swarm. The following calculation represents the modifications of gbest PSO:

$$ve_{ij}(t+1) = iwve_{ij}(t) + ac_1 ra_{1j}(t) [y_{ij}(t) - x_{ij}(t)] + ac_2 ra_{2j}(t) [\hat{y}_{ij}(t) - x_{ij}(t)]$$
(7)

3.5 Fuzzy particle swarm optimization (FPSO)

In order to render the probable solution for feature selection in the churn prediction process, Fuzzy Particle Swarm Optimization (FPSO) algorithm has been applied. The purpose and the advantage of proposed FPSO is by using the fuzzy set theory for adjusting PSO acceleration coefficients adaptively and is thereby able for improve accuracy and efficiency of searches. There is a search space passed through the process concerning the feature selection, and every feature's position is pursued by particle position. By utilizing both individual and social particles experience, the position has revised [36, 37].

The following expression (5) upgrades each particle position for feature selection.

$$x_i(t+1) = x_i(t) + ve_i(t+1)$$
 (5)

where $x_i(t + 1)$ represents the particle position i in the time (t+1) for feature selection in churn prediction,

Through Fuzzy system, PSO's Inertia weight iw and acceleration constants ac₁ and ac₂ of the has adapted. Particle's overall velocity influenced by acceleration constants. Here, the local samples' (population) optimal feature has concluded by the local acceleration constant (ac₁), whereas overall swarm's (datasets) optimal feature depicted by the global acceleration constant (ac_2) . For the reason of converting the inertia weight and acceleration constants, the Fuzzy system has presented; the progress involves the subsequent factor:

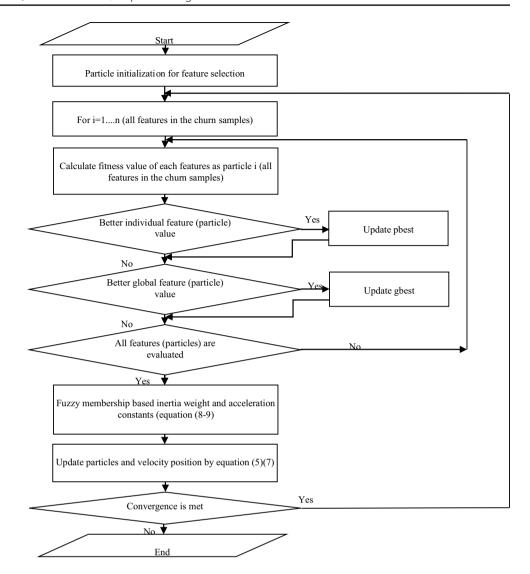
Two inputs: The count of iterations (N) when the optimal fitness is unchanged and original inertia weight value (iw).

Three outputs: the change in inertia weight (ciw) and the change in acceleration constants cac₁ and cac₂.

Figure 2 shows the Flowchart of Fuzzy Particle Swarm Optimization (FPSO) algorithm.

Figure 3 describes regarding the inputs in three membership operations, besides the outputs in fuzzy

Fig. 2 Flowchart of fuzzy particle swarm optimization (FPSO) algorithm



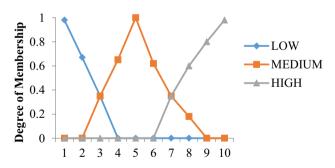


Fig. 3 Triangular membership functions for fuzzy system

techniques. As depicted beneath, the low, medium, and high are being categorized as triangles of left, middle and right. Correspondingly y = trimf(x, params) returns fuzzy membership values computed using following triangular membership function by Eq. (8).

$$f(x;a,b,c) = \begin{cases} 0, x \le a \\ \frac{x-a}{b-q}, a \le x \le b \\ \frac{b-q}{c-b}, b \le x \le c \\ 0, c \le x \end{cases}$$
 (8)

To specify parameters, a, b, and c, use params. Membership values are computed for each input parameters such as Inertia weight iw and acceleration constants ac₁ and ac₂ in PSO algorithm.It is implemented via https://www.mathworks.com/help/fuzzy/trimf.html.

The fuzzy system architecture consisting of two inputs (number of iterations when best fitness does not change (N) and actual inertia weight (iw)) and three outputs (change in inertia weight (ciw)) and acceleration constants cac_1 and cac_2 as shown in Fig. 4.This research utilizes nine fuzzy rules for fetching the new ciw, cac_1 and cac_2 values

Fig. 4 Fuzzy system to adjust the parameters (iw, cac_1 , cac_2)

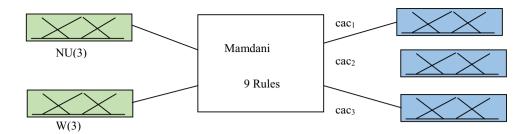


Table 1 Dataset description

Dataset	Instances	Attributes	Attributes name	Target class
telecom churn (cell2cell)	71,047	58	CustomerID, Churn, MonthlyMinutes, Monthly Revenue, TotalRecurringChargeOverageMinutes, RoamingCalls, PercChangeMinutes, Director Assisted Calls, PercChangeRevenues, DroppedCalls, BlockedCalls, CustomerCareCalls, ThreewayCalls, Unanswered Calls, ReceivedCalls, InboundCalls, PeakCallsInOut,Outbound Calls, OffPeakCallsInOut, DroppedBlockedCalls, CallForwardingCalls, CallWaitingCalls, UniqueSubs, ActiveSubs, ServiceArea, Handsets, HandsetModels, Months In Service, CurrentEquipmentDays, AgeHH1, AgeHH2, ChildrenInHH, Handset Refurbished, HandsetWebCapable, TruckOwner, RVOwner, Homeownership, BuysViaMailOrder, RespondsToMailOffers, OptOut-Mailings, NonUSTravelm, Owns Computer, HasCreditCard, RetentionCalls, RetentionOffersAccepted, NewCellphoneUser, NotNew-CellphoneUser, ReferralsMadeBySubscriber, Income Group, OwnsMotorcycle, Adjustment-sToCreditRating, HandsetPrice, MadeCall-ToRetentionTeam, CreditRating, PrizmCode, Occupation, MaritalStatus	Two class classification (Yes –Represent Churn customer, No- Represent Non- Churn Customer)
churn-bigml dataset	3333	21	State, area code, account length, phone number, voice mail plan, international plan, number vmail messages, total day calls, total day minutes, total day charge, total eve calls, total eve minutes, total eve charge, total night calls, total night minutes, total night charge, total intl calls, total intl minutes, total intl charge, customer service calls, churn	Two class classification (T –Represent Churn customer, F- Represent Non- Churn Customer)

Table 2 Fuzz rules to adjust the parameters (iw, ac_1 , ac_2)

Rule number	Input		Output	t	
	N	iw	Ciw	cac ₁	cac ₂
1	1	1	2	1	1
2	1	2	3	2	2
3	1	4	1	4	4
4	2	1	2	1	1
5	2	2	1	3	3
6	2	3	1	3	4
7	3	1	3	1	1
8	3	3	2	3	3
9	3	4	1	2	2

which is illustrated in one fuzzy rule as trials. The medium value of N, high review values corresponds to high ac₁high ac₂ and low ciw. The Table 2 shows the rules acquired by empirical knowledge experimentation. The range of N is in [1, 20], the value for iw is in $0.5 \le iw \le 2.5$, and the values of ac_1 and ac_2 are in $1.0 \le ac_1$, $ac_2 \le 2.0$. The arithmetical explanation for acceleration constants ac₁, ac₂ are expressed with the Eqs. (9–10) discussed below:

$$ac_{1} = \frac{\sum_{i=1}^{ra_{ac_{1}}} \mu_{i}^{ac_{1}} (ac_{1i})}{\sum_{i=1}^{ra_{ac_{1}}} \mu_{i}^{ac_{1}}}$$
(9)

where ac₁ is percent of local acceleration of particlei, ra_{ac}, is number of fuzzy rules initiated to ac₁, ac_{1i} is output of fuzzy rule i for ac₁, and $\mu_i^{ac_1}$ is membership function value of fuzzy rule i for ac₁.

$$ac_{2} = \frac{\sum_{i=1}^{ra_{ac_{2}}} \mu_{i}^{ac_{2}} (ac_{2i})}{\sum_{i=1}^{ra_{ac_{2}}} \mu_{i}^{ac_{2}}}$$
 (10)

where ac_2 is percentage of global acceleration of particle i, ra_{ac_2} is count of fuzzy rules initiated to ac_2 , ac_{2i} is output of fuzzy rule i for ac_2 , and $\mu_i^{ac_2}$ is membership function value of fuzzy rule i for ac_2 . The features chosen from both algorithms of IG and FPSO are widely regarded as foremost features.

3.6 Customer classification and prediction

In the dataset of telecom industry, the customers are classified into two categories. That is to say, the first category includes the customers, those who are stay faithful to their operators and occasionally influenced by competitors namely, non-churn customers; whereas, the second category is called churn customers. The suggested strategy intended to identify the churn customers as well as, the motivation of their departure, besides it conceives retention approaches in order to conquer the issue of migration. With regard to categorising churn datasets, the classifier of Divergence Kernel based Support Vector Machine (DKSVM) has applied in this research, which is suggested for classifying the customers and churn prediction through Kullback-Leibler [38]. Kernel function has applied, to correlate the selected feature datasets of the input consumer churn with the non-churn and churn customers, in which telecom dataset is expected to have proper distribution and, eventually, high-dimensional selected feature space is chosen by this way. The input dataset of telecom with selected features possessed non-churn and churn customers correlated through a new kernel function towards a higher feature space in the DKSVM classifier, and a decision plane is finally created. The ability of DKSVM classifier for generalization has enhanced through maximizing the margin distance and by integrating input dataset of telecom accompanying selected feature to the huge dimensional space, through which the function of nonlinear classification determined and it is considered as a primary advantage of this algorithm. The issue of churn prediction has resolved by this DKSVM classifier through express the Eqs. (11-13).

$$\phi(w, \xi) = \frac{1}{2}(w^*w) + C\left(\sum_{l=1}^{L} \xi_1\right)$$
 (11)

$$\forall \xi_1 \ge 0, \ w^*x_1 + \text{bias} \ge 1 - \xi_1 \text{if} y_1 = +1$$
 (12)

$$w^*x_1 + bias \le -1 + \xi_1 ify_1 = -1$$
 (13)

where 'w' denotes the unknown separation plane, ξ denotes the soft margin, f_l specifies the training telecom dataset with selected features, y_l signifies the respective class of sf_j which is churn and non-chrun. L point out the churn predictiontraining samples, and 'C' denotes the constant. The Lagrange method plays a vital role in churn predictionerror minimization problem which aids in parameter vector identification by up-surging the function given by below Eqs. (14–15)

$$W(\alpha) = \sum_{i=1}^{L} \alpha_1 - \frac{1}{2} \sum_{ij}^{L} \alpha_i \alpha_j y_i y_j K(f_i, f_j)$$
 (14)

$$\sum_{l=1}^{L} \alpha_l y_l = 0, 0 \le \alpha_l \le C \tag{15}$$

 $K(f_i, f_j)$ denotes to the kernel function. Kullback–Leibler [39] has its significance in estimation which is nothing but the probability distribution difference. The Kullback–Leibler divergence from Q to P [39] is given for discrete probability distributions P and Q for f_i , f_i , is as follows (16)

$$D_{KL}(P||Q) = \sum_{f_i, f_i} P(f_i) \log \frac{P(f_i)}{Q(f_j)}$$
 (16)

The Lagrange method is utilized for churn prediction error minimization problem finally for obtaining the new updated kernel and there by achieving work conversion (17)

$$W(\alpha) = \sum_{l=1}^{L} \alpha_1 - \frac{1}{2} \sum_{ij}^{L} \alpha_i \alpha_j y_i y_j ia D_{KL}(P||Q)$$
 (17)

This is the stage for obtaining the customer classification into churn and non- churn by means of MATLAB R2014 tool.

3.7 Customer profiling and relationship

The behavioural information and their relationship are the key factors in partitioning of comprehensive customers' data into groups achieved by Customer cluster concept. The segmentation of telecom data into various groups is attained with the help of Hybrid Kernel Distance based Possibilistic Fuzzy Local Information C-Means (HKD-PFLICM) which also suits partition clustering. The classification is as follows Low, Medium and Risky customers. The integration of Possibilistic-Fuzzy Clustering algorithm (PFCM) [40], and Fuzzy Local Information C-Means (FLICM) algorithm [41] is known as PFLICM. PFLICM helps

in clustering of customers in every segment to the nearest cluster based on the kernel distance. It has advantages such as noise immunity and free of artificial parameters. These combination of PFLICM approaches helps in attaining objective function revealed in Eq. (18):

kernel and Gaussian kernel are the main factors in the computation of kernel distance (ked_{nk}).

Hyperbolic Tangent kernel is termed as Sigmoid Kernels or tanh kernels are different as in Eq. (20),

$$J = \sum_{c=1}^{C} \sum_{n=1}^{N} a u_{cn}^{m} \left(\left| \left| x_{n} - c_{c} \right| \right|_{2}^{2} + G_{cn} \right) + b t_{cn}^{q} \left| \left| x_{n} - c_{c} \right| \right|_{2}^{2} + \sum_{c=1}^{C} \gamma_{c} \sum_{n=1}^{N} \left(1 - t_{cn} \right)^{q}, \ u_{cn} \geq 0, \forall n = 1, \dots N, \ \sum_{c=1}^{C} u_{cn} = 1$$
 (18)

where x_n is a d \times 1 column vector signifying nth telecom data, C is a number of groups being estimated into Low, Medium and Risky customers, c_c is a d \times 1 vector of cth cluster center, weight u_{cn} is a membership value of nth pixel in cth cluster, t_{cn} is typicality value of nthtelecom data in cth cluster, a, b, and α are fixed parameter values used to balance terms of objective function for customer group, and m and q are fixed "fuzzifier" parameters which helps in governing degree of sharing across clusters and degree to which points may be characterized as outliers, respectively. Also,

$$G_{cn} = \sum_{ke \in Ne_n} \frac{1}{ked_{nk} + 1} (1 - u_{ck})^m ||x_k - c_c||_2^2$$
 (19)

In the Eq. (18), ked_{nk} signifies kernel distance between the data indices between x_n and x_k. Ne_n signifies neighborhood around the center data. The hyperbolic tangent

$$ke_1(x_n, x_k) = tanh(v + x_n x_k)$$
(20)

Gaussian kernel is stated as in Eq. (20),

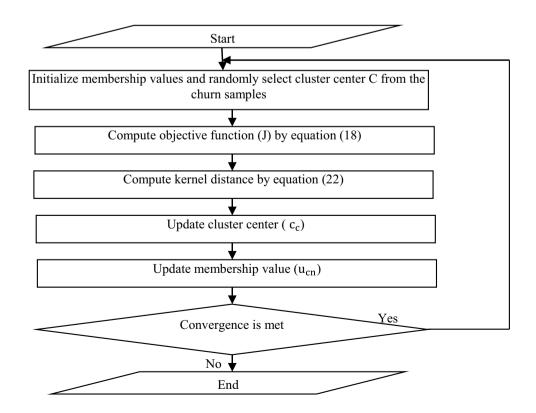
$$ke_{2}(x_{n}, x_{k}) = exp\left(\frac{||x_{n} - x_{k}||^{2}}{2\sigma^{2}}\right)$$
 (21)

The function ked_{nk} helps in combining the kernel outcomes and given by Eq. (22),

$$ked_{nk} = \frac{ke_1(x_n, x_k) + ke_2(x_n, x_k)}{2}$$
 (22)

The overall clustering is achieved for each customer lastly by applying this outcome to Eq. (19). The different pattern and churn customers behaviour are signified by the cluster in each segment. The characteristics of three clusters characterizing a unique collection of behaviour may help in enhancing the decision making scope for

Fig. 5 Flowchart of hybrid kernel distance based possibilistic fuzzy local information C-means (HKD-PFLICM)



churner. Figure 5 shows the overall flowchart of the HKD-PFLICM algorithm.

3.8 Retention strategy uses recommendation system

Appropriate monitoring is required with churn consumers and handles them through the promotion cycle to conclude the retention approaches. Top-down approach has a scarcity with regard to aiming specific consumers, Bottom-up and adapted method is unsuccessful concerning marketing strategy. To build tailored retention practices use the similarity-based strategy. This method is the most familiar definition of a recommender strategy. Based on its expectations and past actions evaluated from clusters, it may assist in proposing acceptable policies or other customized collection of deals to each churn customers. Comparison measure to aim identical consumers relies on the subsequent approaches.

Content-based: This process is focused on past behavior and proposes identical suggestions and products to the same consumer.

Collaborative: This process is focused on past behavior and preference only for those customers who have an identical inclination and comparable behavior.

Hybrid: This process indicates combined approach of both content and collaborative strategies.

The collaborative strategy based approach is proposed for those customers who have identical past behaviours and identical preferences. Every customer is classified into the categories of Low, Medium, and High. These approaches were executed by considering the distance function based on the kernel. In this sense, CRM is associated with personalizing the interaction with consumers and reacting to consumer feedback through company, such as creating a collective management network, expanding front offices to include all staff, collaborators and vendors to communicate with consumers via email, mobile, web sites, contacts, fax, etc. The behaviour of the consumer is tracked and recorded by segmentation to define the patterns and use. From now on, through this pattern and behavior, we render guidelines for only specific customers to suggest.

4 Experiments and results

Numerous trials on the proposed DKSVM strategy have been carried out throughout this section by applying machine learning techniques like, Random Forest (RF), J48 algorithm, Decision Stump [17] on two datasets namely telecom churn (cell2cell) and churn-bigml dataset. The outcomes have been acquired through various machine

 Table 3
 Performance measure of various classification algorithms

 with tenfold cross-validation on cell2cell dataset

Classifiers	Incorrectly classified instances (%)	Correctly classified instances (%)
Decision Stump	20	80
J48	18	82
RF	15	85
DKSVM	10	90

Table 4 Performance measure of various classification algorithms with tenfold cross-validation on churn-bigml dataset

Classifiers	Incorrectly classified instances (%)	Correctly classified instances (%)	
Decision stump	20.11	79.89	
J48	18	82	
RF	13.5430	86.4570	
DKSVM	10.2500	89.7500	

learning strategies, which has been provided in following segments.

4.1 Performance evaluation matrix –churn prediction

The performance of the proposed churn prediction approach has been evaluated through the metrics such as accuracy, precision, f-measure, recall and ROC area. Equation (23) measures accuracy, which recognizes the number of instances that were properly classified.

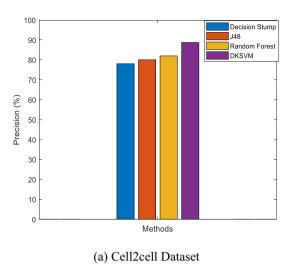
$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$
 (23)

Here "TN" means True Negative, "TP" means True Positive, "FN" means False Negative and "FP" means False Positive. TP Rate is called sensitivity, too. It indicates that part of the data is appropriately marked as positive. The TPR has to be higher for every classifier. Equation (24) is applied to compute True Positive Rate (TPR).

$$TPR = \frac{True Positives}{Actual Positives}$$
 (24)

False Positive Rate (FPR) shows that portion of data is falsely marked as positive. For every classifier the outcome of the FP rate has to be minimal. The estimate is rendered using Eq. (25):

$$FPR = \frac{False \, Positives}{Actual \, Negatives} \tag{25}$$



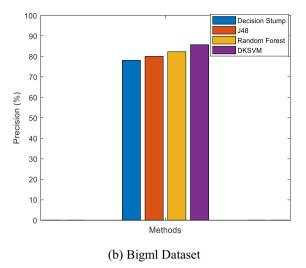


Fig. 6 Precision results comparison vs. classifiers (benchmark datasets)

Table 5 Performance metrics of various classification algorithms on cell2cell dataset

Classifiers	Results (%)					
	Accuracy	Precision	Recall	F-measure	ROC curve	
Decision stump	80	78	79	78	79.12	
J48	82	80	81	81	82.35	
RF	84	82	83	83	84.93	
DKSVM	90	89	90	89	87.65	

Table 6 Performance metrics of various classification algorithms on bigml dataset

Classifiers	Results (%)					
	Accuracy	Precision	Recall	F-measure	ROC curve	
Decision stump	79.89	78.0710	79.4300	77.5430	80.25	
J48	82.0900	80.0321	81.0900	80.7860	83.18	
RF	84.4650	82.3420	83.4560	83.2380	84.87	
DKSVM	87.7500	85.7532	86.7532	86.2503	89.76	

Accuracy, also regarded as positive predictive value (PPV), shows which portion of prediction data is positive. Calculation performed using Eq. (26).

$$Precision = \frac{True \ Positive}{True \ Positive + False \ Positive}$$
 (26)

The recall is further measurement of completeness i.e. the algorithm's real score. It's the chance the process can pick all the appropriate instances. The low recall value

means lots of false negatives. The measurement is done using Eq. (27).

$$Recall = \frac{True Positive}{True Positive + False Negative}$$
(27)

The value of F-measure is a trade-off within categorizing the overall data points accurately and guaranteeing that each class includes points of solely one class. The measurement is performed using Eq. (28).

$$F - measure = 2*\left(\frac{Precision*Recall}{Precision + Recall}\right)$$
 (28)

The ROC field reflects average output against all feasible FP and FN cost ratios. If value of the ROC area is 1.0, then this is a perfect prediction. Equally, value 0.5, 0.6, 0.7, 0.8 and 0.9 signifies random prediction, bad, moderate, good and superior accordingly. The accuracy of overall outcome of the DKSVM algorithm is higher than other prediction algorithms as prediction problem because in both datasets its efficiency is great (See Tables 3 and 4). Tables 3 and 4 reveal the accuracy and building prediction model time for both datasets such as cell2cell and bigml.

Figure 6a and b shows the precision comparison results of four different classifiers in two benchmark datasets. However, the proposed DKSVM classifier has higher precision results of 89.00% and 85.7532% for cell2cell dataset and bigml dataset respectively (See Tables 5 and 6). The proposed DKSVM gives 7.6822%, 5.7211% and 3.4122% higher precision results when compared to Decision Stump, J48 and RF classifiers respectively in Bigml dataset. Decision stump gives worst precision results for CELL-2CELL dataset when compared to other methods. RF and

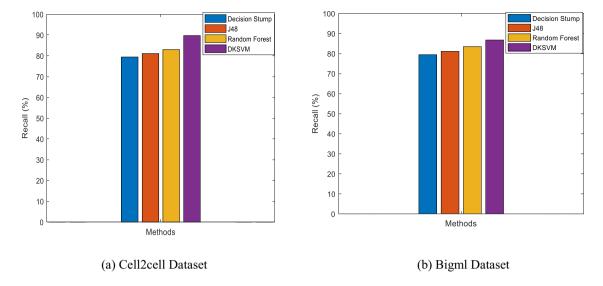


Fig. 7 Recall results comparisonvs. classifiers (benchmark datasets)

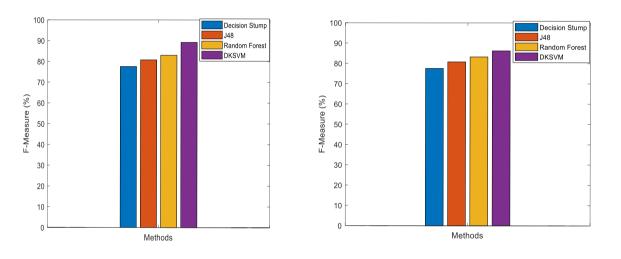


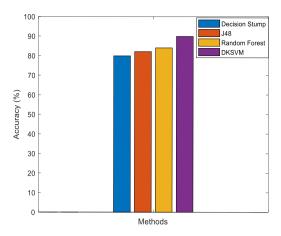
Fig. 8 F-measure results comparisonvs. classifiers (benchmark datasets)

proposed classifiers have high precision to predict churners correctly in both datasets as shown in Tables 5 and 6.

Figure 7a and b shows the performance comparison results of the two benchmark datasets via recall among four different classifiers. However, the proposed DKSVM classifier has higher recall results of 90% and 86.7532% for cell2cell dataset and bigml dataset respectively (See Tables 5 and 6). It means that this algorithm found maximum number of false positives in dataset and it can correctly identify non-chrun customers. This indicates that they are good classifiers (DSVM and J48) for prediction. The proposed DKSVM gives 7.3232%, 5.6632% and 3.2972% higher when compared to Decision Stump, J48

and RF classifiers respectively in second dataset. When compared to existing methods Decision Stump and RF gives worst recall results, since tree based models have incorrectly predicts the false values.

Figure 8a and b shows the four different classifiers in two benchmark datasets with respect to f-measure. Proposed DKSVM classifier has higher f-measure results of 89% and 86.2503% for CELL2CELL dataset and BIGML dataset respectively (See Tables 5 and 6). The proposed DKSVM gives 8.7073%, 5.4643% and 3.0123% higher f-measure results when compared to Decision Stump, J48 and RF classifiers respectively for Bigml dataset. When compared to existing methods, Decision Stump gives



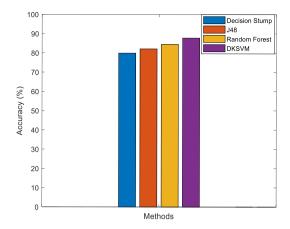


Fig. 9 Accuracy results comparison vs. classifiers (benchmark datasets)

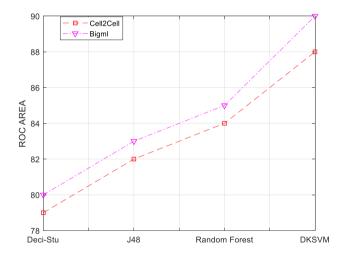


Fig. 10 ROC area curve. vs. classifiers (benchmark datasets)

worst f-measure results of 73.9945% for CELL2CELL dataset. Other methods having higher F-measure in both datasets as shown in Tables 5 and 6.

Figure 9a and b shows the performance comparison results of the two benchmark datasets with accuracy results comparison among four different classifiers. The proposed DKSVM classifier has higher accuracy results of 90% and 87.75% for cell2cell dataset and bigml dataset respectively (See Tables 5 and 6). It means that this algorithm found the maximum number of true positives in the dataset and it can correctly identify the churn customers. The proposed DKSVM gives 7.86%, 5.66% and 3.285% higher accuracy results when compared to Decision Stump, J48 and RF classifiers respectively. Random Forest and proposed DKSVM algorithms have performed very well as compared to all other algorithms by having 87.65% and 89.76% ROC area under the curve for first and second dataset. DKSVM is an outstanding classifier for prediction of instances. According to the ROC value scale discussed earlier, RF and DKSVM also performed better as shown in the Fig. 10 (see Tables 5 and 6).

4.2 Performance evaluation matrix -clustering

The Customer Cluster is employed to segment the entire consumer details into categories on the basis of their activities and interaction detail. In order to measure the results of the clustering methods the following metrics such as accuracy and error is measured in the telecom churn (cell2cell) and churn-bigml datasets (Table 7). The results

Table 7 Clustering metrics comparison vs. clustering algorithms

Metrics/methods	K-means	Flexible K-Medoids	FLICM	EWFLICM	PFLICM	HKD-PFLICM
Dataset	cell2cell da	taset				
Accuracy (%)	76.51	80.498	90.41	92.56	94.11	96.48
Error(%)	23.49	19.502	9.59	7.44	5.89	3.52
Dataset	bigml data	bigml dataset				
Accuracy (%)	81.19	87.43	91.25	93.58	95.41	97.17
Error(%)	18.81	12.57	8.75	6.42	4.59	2.83

 Table 8
 Threshold values of churn customers (churn-bigml dataset)

Calls	Risky (Cluster 1)	Medium(Cluster 2)	Low(Cluster 3)
Total calls behav	ior in each cluste	r	
Total day calls	5	60	31
Total eve calls	32	55	9
Total night calls	14	48	34
Total intl calls	46	43	7
Total minutes be	havior in each clu	ıster	
Minutes			
Total day minutes	44	20	66
Total eve minutes	14	30	52
Total night minutes	30	15	51
Total intl minutes	25	38	33
Total charge beh	avior in each clus	ter	
Charge			
Total day charge	24.6	34.2	37.91
Total eve charge	18.3	21.86	26.8
Total night charge	12.93	10.8	8.36
Total intl charge	3.89	3.7	2.35

Table 9 Threshold values of churn customers (Churn- Cell2cell Dataset)

Calls	Risky (Cluster 1)	Medium(Cluster 2)	Low(Cluster 3)					
Communication ca	Communication calls behavior in each cluster							
Peak call	145	113	100					
Offpeak call	80	56	37					
Received call	10	66	97					
Minutes behavior i	n each cluste	r						
Minutes								
Month min	43	8	13					
Overage min	76	62	27					
PercChange min	13	27	5					
Uncommunication calls behavior in each cluster								
Dropped call	137	56	43					
Blocked call	144	66	25					
Unanswered call	43	25	110					

of the proposed HKD-PFLICM are compared to K-means, Flexible K-Medoids, PFLICM, Fuzzy Local Information C-Means (FLICM), and Entropy Weighting LICM known as

EWFLICM.Accuracy measurements demonstrate the level of similarities between one cluster (i.e., set of clusters) is to another cluster (ground-truth). Error metrics depict the level of non-similarities between one cluster (i.e., set of clusters) to another (ground-truth) clustering. Customer cluster is used for partitioning complete customers' data into groups based on their behavior information and their relationship. For measuring results of clustering methods following metrics such as accuracy and error is measured in the telecom churn (cell2cell) and churn-bigml datasets (Table 7).

The results of the proposed HKD-PFLICM are compared to K-means, Flexible K-Medoids, PFLICM, Fuzzy Local Information C- Means (FLICM), and Entropy Weighting LICM known as EWFLICM. Accuracy metrics indicate how much one clustering (i.e., set of clusters) is similar to another (ground-truth) clustering. Error metrics indicate how much one clustering (i.e., set of clusters) is non-similar to another (ground-truth) clustering (Table 8).

Clustering behavior of churner with three major attributes such as communication calls, uncommunication calls, minutes having forms three clusters such as Low (Cluster 3), Medium (Cluster 2) and Risky (Cluster 1) for CELL2CELL dataset. Those cluster threshold values are discussed in the Table 9.

Figure 11 shows the clustering accuracy and error results comparison among various clustering methods with two benchmark datasets. The proposed clustering algorithm gives higher accuracy results of 94.11% and 95.41% for cell2cell dataset and bigml dataset. Since the proposed work, instead of considering the distance between the customer behaviour kernel functions is used which increases the clustering results than the other methods.

From this pattern and behaviour, make rules for the recommendation of only similar customers in the future. Figures 12, 13 and 14 show the behaviour of churner with three major attributes such as TOTAL CALLS, TOTAL MIN-UTES, TOTAL CHARGE having forms three clusters such as Low (Cluster 3), Medium (Cluster 2) and Risky (Cluster 1). Those cluster threshold values are discussed in the Table 8.

5 Conclusion and future work

Customer Relationship Management (CRM) always have the major concern of churn prediction, in order to keep hold of, as well as to improve the number of precious customers through recognizing an identical set of new customers and furnishing them a considerable discounts, services and offers. Through this paper, an efficient strategy of customer churn prediction is presented, which is an optimal choice to support telecom service providers to

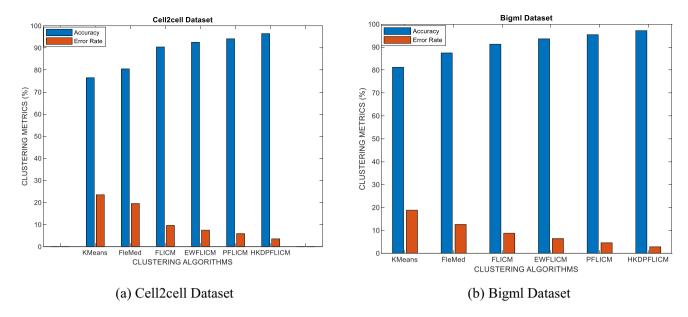


Fig. 11 Accuracy and error results comparison between clustering methods

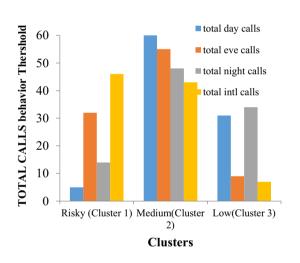


Fig. 12 Total_Calls behavior in each cluster (churn-bigml dataset)

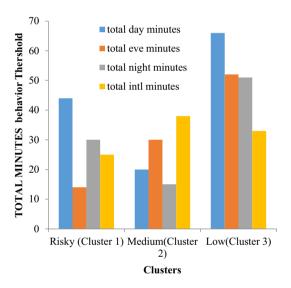


Fig. 13 Total minutes behavior in each cluster (churn-bigml dataset)

identify consumers, as they apparently dealt with churn. To perform this task, some data mining steps have been taken to account, such as noise removal, feature selection, customer classification and prediction. This proposed work consists of the following steps: feature selection, customer classification and prediction and finally customer profiling. Information Gain (IG) and Fuzzy Particle Swarm Optimization (FPSO) are proposed for feature selection. By employing inertia weight and acceleration coefficients, the world's optimum version of PSO has amended, concerning enrichment namely, FPSO. To process the categorization of Churn and Non-Churn patrons, the classifier

of Divergence Kernel-based Support Vector Machine (DKSVM) has been applied through the metric focused on Kullback-Leibler. Eventually, the characteristics of customers have procured from customer profiling and aggregated as clusters by enforcing the clustering method of Hybrid Kernel Distance-based Possibilistic Fuzzy Local Information C-Means (HKD-PFLICM) concerning the CRM to maintain the customers. The similarities among overall customers have been observed from the analysis, through which

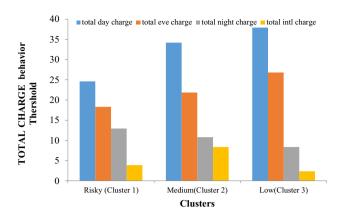


Fig. 14 Total charge behavior in each cluster (churn-bigml dataset)

the concern's productive potential has enriched, thus the concern eventually reaches the competent marketing promotions. The implementation outcomes represents that, the proposed DKSVM churn approach accompanied by classifier approaches is capable of delivering efficient performance. For instance, the proposed clustering algorithm gives higher accuracy results of 94.11% and 95.41% for cell2cell dataset and bigmldataset. Since the proposed work uses the kernel functions instead of considering the distance between the customer behaviour, the clustering results are observed to be better than the other existing methods. In future, the additional exploration of this approach will focus on the learning strategies of eager learning and lazy learning for optimum churn prediction. In addition, this research has augmented to scrutinize the revising character pattern of churn customers by enforcing the strategies of Artificial Intelligence (AI), which are intended to analyze the trend and prophecy. Apart from SVM, some other prediction approaches like Genetic Algorithms (GAs), Neural Networks (NNs), etc. can be used. Moreover, various other churn prediction datasets could be used and its results could be analyzed in future.

Funding This study received no specific financial support.

Availability of data and material Not Applicable.

Code availability Granger causality testing.

Declarations

Conflict of interest The authors declare that there are no conflicts of interest.

Ethics approval There are no waivers for this project. (Not Applicable)

Consent to participate Not applicable.

Consent for publication Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

- Babu S, Ananthanarayanan DN, Ramesh V (2014) A survey on factors impacting churn in telecommunication using datamining techniques. Int J Eng Res Technol (IJERT) 3(3):1745–1748
- Lu N, Lin H, Lu J, Zhang G (2012) A customer churn prediction model in telecom industry using boosting. IEEE Trans Industr Inf 10(2):1659–1665
- Wei CP, Chiu IT (2002) Turning telecommunications call details to churn prediction: a data mining approach. Expert Syst Appl 23(2):103–112
- Qureshi SA, Rehman AS, Qamar AM, Kamal A, Rehman A (2013) Telecommunication subscribers' churn prediction model using machine learning. In: Eighth international conference on digital information management (ICDIM 2013), pp 131–136
- Ascarza E, Iyengar R, Schleicher M (2016) The perils of proactive churn prevention using plan recommendations: evidence from a field experiment. J Mark Res 53(1):46–60
- Idris A, Khan A (2012) Customer churn prediction for telecommunication: employing various various features selection techniques and tree based ensemble classifiers. In: 2012 15th international multitopic conference (INMIC), pp 23–27
- Kaur M, Singh K, Sharma N (2013) Data mining as a tool to predict the churn behaviour among Indian bank customers. Int J Recent Innov Trends Comput Commun 1(9):720–725
- 8. Popović D, Bašić BD (2009) Churn prediction model in retail banking using fuzzy C-means algorithm. Informatica 33(2):1–6
- Owczarczuk M (2010) Churn models for prepaid customers in the cellular telecommunication industry using large data marts. Expert Syst Appl 37(6):4710–4712
- Kim N, Jung KH, Kim YS, Lee J (2012) Uniformly subsampled ensemble (USE) for churn management: theory and implementation. Expert Syst Appl 39(15):11839–11845
- 11. Anil Kumar D, Ravi V (2008) Predicting credit card customer churn in banks using data mining. Int J Data Anal Tech Strateg 1(1):4–28
- Hadden J, Tiwari A, Roy R, Ruta D (2007) Computer assisted customer churn management: state-of-the-art and future trends. Comput Oper Res 34(10):2902–2917
- Kim HS, Yoon CH (2004) Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market. Telecommun Policy 28(9–10):751–765
- Huang Y, Kechadi T (2013) An effective hybrid learning system for telecommunication churn prediction. Expert Syst Appl 40(14):5635–5647

- Sharma A, Panigrahi D, Kumar P (2013) A neural network based approach for predicting customer churn in cellular network services. Int J Comput Appl 27(11):26–31
- Ali ÖG, Arıtürk U (2014) Dynamic churn prediction framework with more effective use of rare event data: the case of private banking. Expert Syst Appl 41(17):7889–7903
- Ullah I, Raza B, Malik AK, Imran M, Islam SU, Kim SW (2019)
 A churn prediction model using random forest: analysis of machine learning techniques for churn prediction and factor identification in telecom sector. IEEE Access 7:60134–60149
- 18. Tsai CF, Lu YH (2009) Customer churn prediction by hybrid neural networks. Expert Syst Appl 36(10):12547–12553
- Jahromi AT, Moeini M, Akbari I, Akbarzadeh A (2010) A dualstep multi-algorithm approach for churn prediction in pre-paid telecommunications service providers. J Innov Sustain RISUS 1(2):2179–3565
- Verbeke W, Martens D, Mues C, Baesens B (2011) Building comprehensible customer churn prediction models with advanced rule induction techniques. Expert Syst Appl 38(3):2354–2364
- Zhang X, Zhu J, Xu S, Wan Y (2012) Predicting customer churn through interpersonal influence. Knowl-Based Syst 28:97–104
- 22. Mishra A, Reddy US (2017) A novel approach for churn prediction using deep learning. In: Proceedings of the IEEE international conferences computing intelligent computer research, pp 1–4
- Amin A, Anwar S, Adnan A, Nawaz M, Alawfi K, Hussain A, Huang K (2017) Customer churn prediction in the telecommunication sector using a rough set approach. Neurocomputing 237:242–254
- Ravi S, Balakrishnan PA (2011) Genetic algorithm based temperature controller for plastic extrusion system. ICGST ACSE J 11(1):1–8
- 25. Mitrović S, Baesens B, Lemahieu W, De Weerdt J (2018) On the operational efficiency of different feature types for telco Churn prediction. Eur J Oper Res 267(3):1141–1155
- De Caigny A, Coussement K, De Bock KW (2018) A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. Eur J Oper Res 269(2):760–772
- Stripling E, vandenBroucke, S., Antonio, K., Baesens, B., & Snoeck, M. (2018) Profit maximizing logistic model for customer churn prediction using genetic algorithms. Swarm Evol Comput 40:116–130

- 28. Agrawal S, Das A, Gaikwad A, Dhage S (2018) Customer churn prediction modelling based on behavioural patterns analysis using deep learning. In: 2018 international conference on smart computing and electronic enterprise (ICSCEE), pp 1–6. IEEE.
- Amin A, Shah B, Khattak AM, Moreira FJL, Ali G, Rocha Á, Anwar S (2019) Cross-company customer churn prediction in telecommunication: a comparison of data transformation methods. Int J Inf Manage 46:304–319
- Tian DP, Li NQ (2009) Fuzzy particle swarm optimization algorithm. In: IEEE international joint conference on artificial intelligence, 263–267
- 31. Tian D, Shi Z (2018) MPSO: modified particle swarm optimization and its applications. Swarm Evol Comput 41:49–68
- 32. Neshat M (2013) FAIPSO: fuzzy adaptive informed particle swarm optimization. Neural Comput Appl 23(1):95–116
- Yang YS, Chan FH, Lam FK, Nguyen H (1997) A new fuzzy classifier with triangular membership functions. In: Proceedings of international conference on neural networks (ICNN'97), vol 1, pp 479–484). IEEE
- 34. Yu HF, Lo HY, Hsieh HP, Lou JK, McKenzie TG, Chou JW, Weng JY (2010) Feature engineering and classifier ensemble for KDD cup 2010. In: KDD cup, pp 1–16
- Zhao L, Gao Q, Dong X, Dong A, Dong X (2017) K-local maximum margin feature extraction algorithm for churn prediction in telecom. Clust Comput 20(2):1401–1409
- Abdalla TY, Abed AA, Ahmed AA (2017) Mobile robot navigation using PSO-optimized fuzzy artificial potential field with fuzzy control. J Intell Fuzzy Syst 32(6):3893–3908
- Hashim HA, El-Ferik S, Abido MA (2015) A fuzzy logic feedback filter design tuned with PSO for L1 adaptive controller. Expert Syst Appl 42(23):9077–9085
- 38. Joyce JM (2011) Kullback-leibler divergence. In: International encyclopedia of statistical science, pp 720–722
- Polani D (2013) Kullback-leibler divergence. In: Encyclopedia of systems biology, pp 1087–108
- Pal NR, Pal K, Keller JM, Bezdek JC (2005) A possibilistic fuzzy c-means clustering algorithm. IEEE Trans Fuzzy Syst 13(4):517–530
- 41. Krinidis S, Chatzis V (2010) A robust fuzzy local information C-means clustering algorithm. IEEE Trans Image Process 19(5):1328–1337

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.