

AN EFFECTIVE MACHINE LEARNING MODEL FOR CUSTOMER ATTRITION PREDICTION IN MOTOR INSURANCE USING GWO-KELM ALGORITHM

Deepthi Das¹, Raju Ramakrishna Gondkar²

¹ Research Scholar, Department of Computer Science, CMR University, (India)
Associate Professor, Department of Computer Science, CHRIST (Deemed to be University),(India)

² Professor, Department of Computer Science, CMR University,(India)

ABSTRACT:

The prediction of customers' churn is a challenging task in different industrial sectors, in which the motor insurance industry is one of the well-known industries. Due to the incessant upgradation done in the insurance policies, the retention process of customers plays a significant role for the concern. The main objective of this study is to predict the behaviors of the customers and to classify the churners and non-churners at an earlier stage. The Motor Insurance sector dataset consists of 20,000 records with 37 attributes collected from the machine learning industry. The missing values of the records are analyzed and explored via Expectation Maximization algorithm that

categorizes the collected data based on the policy renewals. Then, the behavior of the customers are also investigated, so as to ease the construction process training classifiers. With the help of Naive bayes algorithm, the behaviors of the customers on the upgraded policies are examined. Depending on the dependency rate of each variable, a hybrid GWO-KELM algorithm is introduced to classify the churners and non-churners by exploring the optimal feature analysis. Experimental results have proved the efficiency of the hybrid algorithm in terms of 95% prediction accuracy; 97% precision; 91% recall & 94% F-score.

Keywords - Customer chum, Customer

retention, Extreme Learning Machine (ELM), Grey Wolf Optimizer (GWO), Motor Insurance

1. Introduction

Motor Industry is one of the well-known and leading industries that make use of data mining techniques [1]. It is mainly on the fact that a vast amount of data with incessant increased rate of customers relies on the insurance companies. It poses a challenging scenario for the researchers as well as industrialists. Since the customers have the ability to switch over among the available providers, the prediction of the churn rate in the motor insurance industry is not performing well in the aspects of customer's requirements. Therefore, it is addressed in terms of churn prediction using machine learning techniques [2-4]. The insurance industry is extensive and changes rapidly according to the global market trends. Developing a trustable and reliable Customer Relationship Management (CRM) [5-7] in the motor industry is an interesting study area which stimulates to deal with the real-time challenges. The prediction of churn in the motor insurance industry is a significant research area for the fault recognition and classification tasks. The rise of churners in several businesses are due to

the inefficient tools development that fails to predict the risk of customer churn.

Generally, each data item in the communication environment consists of churn that can be identified with the help of background information. The churn of a customer is defined as the process of ceasing the customer connection with the company [8]. In the context of an online business environment, customers turn out to be churned, when the customer communication with the service time has elapsed. This could bring revenue lost and the deprived marketing costs due to the improper customer communication. Therefore, customer churn should be minimized which is the ultimate aim for all online business environments [9]. The main challenges in the customer churn prediction systems are: i) do not associate with the business objective and ii) relying on feedback loops. These are the two reasons that have increased the count of churners in the business environment. Customers who are willing to churn should be excluded from the campaign, so as to preserve the resources.

The rest of the paper is organized as follows: Section II presents the Related work; Section III presents the Research Methodology; Section IV presents the

Experimental Results and Analysis and Section V presents the Conclusion.

2. Related work

This section presents the reviews of existing machine learning techniques designed for churn predictions. The author in [10] has discussed the market segmentation using hierarchical self-organizing maps. It was explored in a real-time dataset of market segmentation in Taiwan. The classification expected error and the associated input vectors were organized in a hierarchical way. This helps to build a customer map that effectively brings out the stage of customers under the churning stage. However, the clustering process develops a class imbalance issues over the multi-subscription customer data. Detection of potential churners at the earlier stage is the complicating task due to the invasion of new subscription services [11]. It was explored on the European pay TV company where a high rate of churners were observed. With the help of markov chain concepts, the reasons behind the customer attrition were found. These were then fed into RF classifiers, so as to effectively classify the churners. Depending on the contact variables, the technical problems were resolved. The author in [12] has discussed

the key issues and the future directions of customer retention management. The role of customer retention was to increase the firm value by deducing the rate of subscription based firms. By estimating the social-connectivity factors, the churns were predicted. Campaign management was the most composed strategy that introduced the win-back scenario on all customer retention. All the data were leveraged based on the retention values.

Intelligent data analysis and their approaches were studied to resolve the business issues [13]. Due to the enhancement in the globalization process, an in-depth analysis of churn in the sense of customer continuity management. It has stated the use and the merits of the prediction models with the business solution. Some recommendations were given according to the scope of industry. The study has concluded that the scope of deep learning and ensemble learning might even predict faster than the machine learning techniques. In [14], the author has introduced a random forest algorithm along with the use of regression analysis, in order to predict the profitability and retention of the customers. It was explored on the 100,000 customers taken from the warehouse of

European financial services company. The RF algorithm was applied on two aspects, one is to classify the binary data and second is to leverage the linear dependent variables. Compared to the logistic regression models, the behavior analysis of the customers was done by exploring the buying variables and the profitability variables. Though it has suggested to increase the profitability, however, the multi-objective based behavior patterns can also be focussed. The author in [15] has suggested the scope of random forest algorithms in order to enhance the uplift modeling on customer retention cases. Here, a customer retention model was designed to the customers who were actively engaged. By the use of a random forest algorithm, the customer attrition rate was leveraged. Strong empirical evidence was suggested to retain the customers for long-term. Though it has easily predicted the retentional customers, the effects of attrition that leads to customer retention are not recommended.

Biological inspired approaches were introduced to predict the churn, so as to yield the optimal performance. PSO and its variants [16] were introduced to effectively preprocess the data. With the help of optimized feature selection process and the

simulated annealing process, the rate of false positivity and the choice of features complexity were significantly reduced. Though it has provided an optimal strategy to predict the churn, yet the time complexity and termination criteria are high which brings a heavy computational effort. In [17], a random rough subspace based neural ensemble learning process was developed to improve the performance of the neural network classifiers. It was combined with the set of feature reductions process, so as to increase the effectiveness and the efficiency of the ensemble strategies. The data proportion of the classifier misleads the data relationship. One-class SVM with undersampling data [18] was studied to improve the churn prediction along with insurance fraud. It is analyzed on the Automobile Insurance fraud dataset and Credit card customer churn dataset by significantly improving the AUC performance. Compared to other ML techniques, the OC-SVM has predicted the churners with the least computational steps. However, the class imbalance and false positive rate are higher, when the sensitivity learning values increases.

In the context of B2B e-commerce companies, the SVM [19] was employed to

enhance the AUC parameter-selection technique. In the perspective of churn prediction, a hyperplane in high dimensional space was constructed using SVM that optimally discriminates between the churners and non-churners. It also marginalized the different hyperplanes of the data classes. Though this model has effectively predicted the churn and increased the revenue costs, the subscription renewal service is not focussed. Subscription services are one of the challenging tasks in dealing churn prediction. Initially, a churn model was designed on the profile behavior analysis. Grid search is performed on the obtained noise marketing data. Skewing data is likely to increase the support vectors. Thus, an optimized learning parameters with SVM has achieved 4.492 decile score than the logistic regression and random forests. Some variable groups are dynamic over data. In [20], Support Vector Machine (SVM) was studied to analyze the customer churns in the case of credit cards. In order to increase the churn prediction accuracy, the learning unit in SVM was redefining the feature space. It has achieved 94.7% of training accuracy and 82.4 % of testing accuracy, compared to the ANN and BPN. However, the various parameters related to customer satisfaction are not considered.

From the conducted reviews, the research problems pertaining in enhancing the CRM in the motor insurance industry are presented. In the angle of machine learning techniques, the churns are predicted based on the customer association with each subscription service. The problems can be viewed in three forms, namely, testing phase, training phase and the prediction phase. In training phase viewpoint, the labels are assigned based on the available churn data, which develops a class imbalance during the prediction phases leading to the degradation of classifying accuracy. In prediction phase viewpoint, the possibility of predicting the churners based on the retention policies which confuse the classifier in the context of multi-dimensional data variables.

3. Research Methodology

This section presents the proposed classification scheme using GWO-KELM. The proposed phases are:

3.1 Data preprocessing:

It is the first step that cleans the collected data by introducing a sampling process. It contains two steps, namely, E-Step and M-Step. The aim of these two steps are to eliminate the irrelevant data to develop

better prediction model systems. Here, the concept of Expectation-Maximization algorithm is employed to estimate the likelihood behavior on the two sets of random variables.

- a) E-step: In this step, the unlabeled class data is estimated by deriving the expectation on the missing value. The probability distribution is derived over all the considered attributes.
- b) M-Step: It derives the maximum likelihood for all considered attributes.

The followings are the steps involved in preprocessing the acquired data:

- i) Let the data be labeled as D; Missing values as M and δ be the unknown parameter. It is mathematically expressed as:

$$\dots(1)$$

be the expression of likelihood function:

$$L(\delta \vee D) \in \{ \alpha p(D \vee \delta) \dots(2)$$

- ii) For the considered variables, D, M and δ , the calculation of maximum likelihood is done by the marginal likelihood on the data. The estimation of the maximum likelihood

is given as:

$$L(\delta; D) = p(D \vee \delta) = \sum_{\epsilon} p(D, Z) \vee \delta \dots(3)$$

- iii) Then, the expected value of the log likelihood function is given as:

$$Q(\delta \vee \delta \wedge (t)) = EM \vee x_i \delta \wedge (t) [\log L(\delta; D, M)] \dots(4)$$

- iv) Discovering the parameters that maximizes the data is given as:

$$\delta \wedge (t + 1) = \text{argmax} Q(\delta \vee \delta \wedge (t)) \dots(5)$$

The above steps (i) to (iv) are iteratively applied on the acquired data, until the data is eligible to develop prediction model systems.

3.2 Feature extraction & Feature selection process:

From the preprocessed data, the behaviors of the customers are analyzed, so as to select the relevant features for constructing efficient classifiers. Here, the Naive Bayes technique is employed to model the behavior of the customers. Let us consider a set of instances with a vector of features as, $[a_1, a_2, \dots, a_n]$ and the bayes theorem for the

observed data is expressed as:

$$P(b_q \vee a_1, a_2, \dots, a_n) = \frac{P(b_q)P(a_1, a_2, \dots, a_n \vee b_q)}{P(a_1, a_2, \dots, a_n)} \dots(6)$$

Where,

The class probability is represented as b_q which is mentioned in training data.

While observing the data, the probability of the class is represented as, P

Then, the data instances that has the highest probability is given as:

$$b_{MAP} = \arg \max_{y_j \in C} P(b_q)P(a_1, a_2, \dots, a_n \vee b_q) \dots(7)$$

With the help of the above classification rule, the probability is estimated over the training data. Depending on the obtained probability rate, the relevant features of the churn prediction are discovered.

3.3 Classification & Prediction process:

In this step, Grey Wolf Optimization (GWO)- Kernel Extreme Learning Machine (KELM) explains the process of classification and prediction by considering the optimal selection of attributes. The selected features are then optimized using

GWO and then, it is used for classification purposes by introducing the KELM model. The kernel based extreme learning machine consists of three layers namely, input I, output O and hidden H. Initially, the connection weights to the inputs are arbitrarily selected for the input and hidden layer. During the training process, the threshold of weights is being adjusted in the networks. An unique optimal solution is obtained by triggering the hidden layers which is done by GWO algorithm. The observed output between the hidden layers is computed by the activation function which is expressed as:

$$\sum_{i=1}^L \beta_i g_i(x_j) = \sum_{i=1}^L \beta_i g_i(w_i \cdot x_j + b_i) = t_j \dots(8)$$

where, the input nodes with hidden nodes L are optimized using GWO. Inspired from the hunting behavior of the wolf, the GWO algorithm is designed. The proposed steps of the GWO algorithm are:

- a) It consists of four groups, namely, Alpha (α); Beta (β); Delta (δ) and omega (ω) in which the 1st three groups determines the fitness criteria and the 4th group determines the optimal position of the wolf.
- b) The position of the wolf is

determined from below two equations:

$$\underline{D} = \underline{X_p}(t) \cdot \underline{C} - \underline{X}(t) \quad \text{.....(9)}$$

$$\text{.....(10)}$$

Where,

The difference between the position and predator is represented as, \underline{D}

The current iteration is given as t.

The position of the prey is given as $\underline{x_p}$.

The position of the wolf is given as \underline{x}

Above all, the vectors A and C are the most important parameters that determine the accuracy of the optimization process:

$$\underline{A} = r1a.2a \quad \text{.....(11)}$$

$$\underline{C} = r2.2 \quad \text{.....(12)}$$

Where,

The vectors r1 and r2 are the random values ranging from 0 to 1.

According to the position of the class , the position of the features are updated

iteratively in the position (X, Y).

The pseudocode of the GWO algorithm:

Declaring the population P_i where, $i = 1, 2, \dots, n$

Initializing the a, A and C.

The fitness criteria of the each searching process is estimated as,

$P_\alpha \rightarrow$ 1st searching agent

$P_\beta \rightarrow$ 2nd searching agent

$P_\delta \rightarrow$ 3rd searching agent

While (iter < stopping criteria)

For each searching agent

Alter the present the position of the searching agent by below equation,

$$\underline{P}(t + 1) =$$

End for

Alter A, a & C

Again the fitness value is computed,

Altering $P_\alpha, P_\beta, P_\delta$

iter= iter + 1

End while

Return P_α

Finally, the A and C are the regularization coefficient that triggers the position of the class according to the

optimized features. Mostly, the small weights are taken for the training procedure and thus, a minimized error function is achieved. Depending on the settings of the parameters, the stability of the churners behaviors are predicted with the least computational steps.

4. Experimental results and analysis

This section presents the experimental setup parameters and the performance measures considered in this study. The achieved results are discussed in a detailed manner. The insurance sector database collected from the machine learning repository is used for experimental purposes. The dataset is collected over a period of a year and a half that incorporate the details of each customer recorded throughout the specified time period. It is tested on Python, a high-level programming language. It consists of 37 attributes and 20,000 records, in which only few of them are taken for constructing the training classifier. The hybrid algorithm is compared with the classical KELM. The performance metrics analyzed in this study are accuracy, precision, recall and F-score which is displayed via confusion matrix in Fig.1.

Premises:

Labels: Churners and Non- Churners

True positive (TP)- No. Of samples correctly identified as abnormal

False positive (FP)- No. Of samples incorrectly identified as abnormal

True Negative (TN)- No. Of samples correctly identified as normal

False Negative (FN)- No. Of samples incorrectly identified as normal.

a) Recall: It is defined as the proportion of real positive cases that are correctly predicted positive. It is given as:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad \dots(13)$$

b) Precision: It is defined as the proportion of predicted positives that are correctly real positives. It is given as:

$$\text{Precision} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad \dots(14)$$

c) Prediction Accuracy: It is the most intutional performance that measures the correctly predicted observation to the total observations. It defines the ability of distinguishing normal and abnormal cases. It is given as:

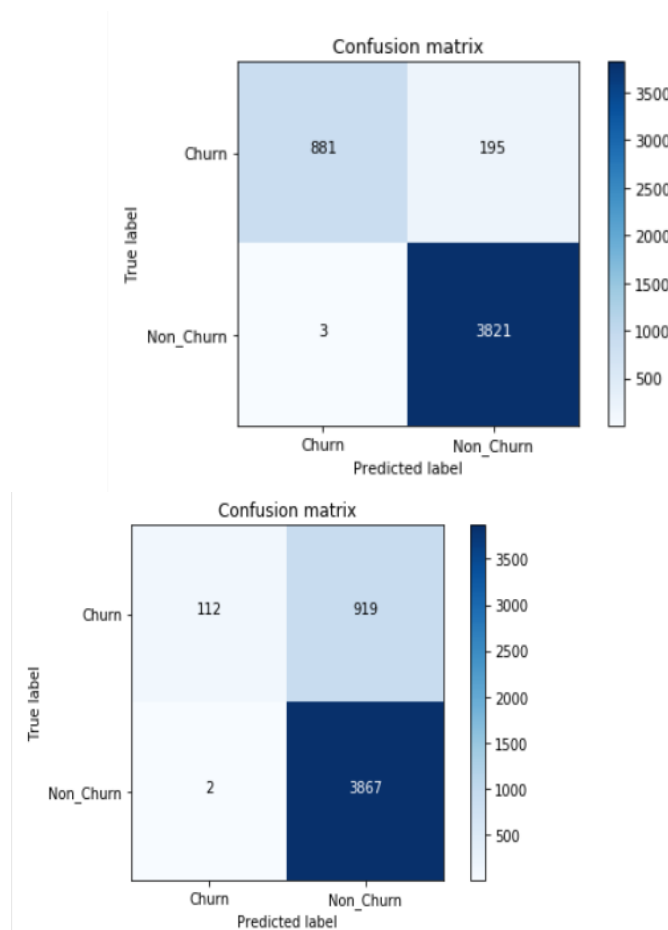
$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad \dots(15)$$

d) F-score: It conveys the balance between precision and recall which is given in the

below equation:

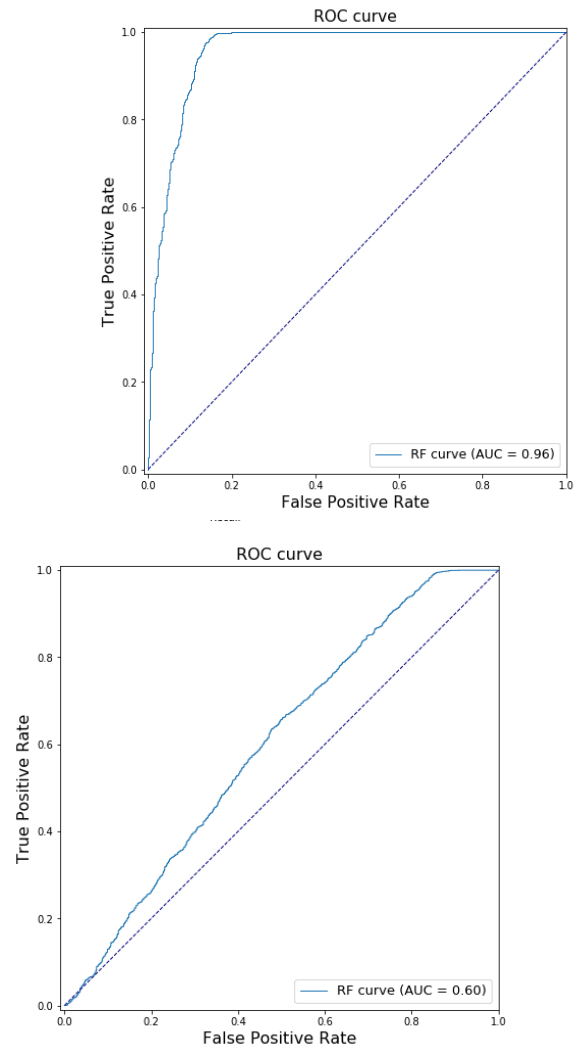
$$F\text{-score} = \frac{2 * ((\text{precision} * \text{recall}) / (\text{precision} + \text{recall}))}{\dots(16)}$$

Since the proposed objectives consist of 2 classes, a confusion matrix is displayed as follows:



(a) Proposed
(b) Existing

Figure1. Confusion matrix between the proposed and existing algorithm

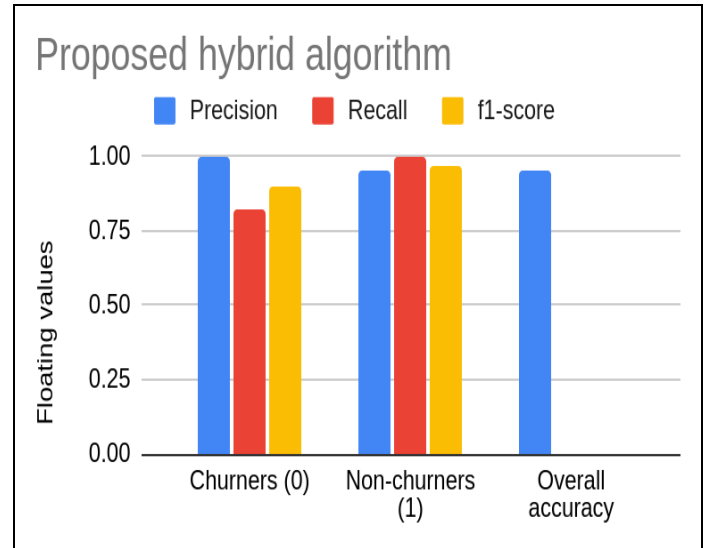


(a) Proposed
(b) Existing

Figure2. ROC analysis between the proposed and existing algorithm

Table 1. Achieved results - Proposed and Existing

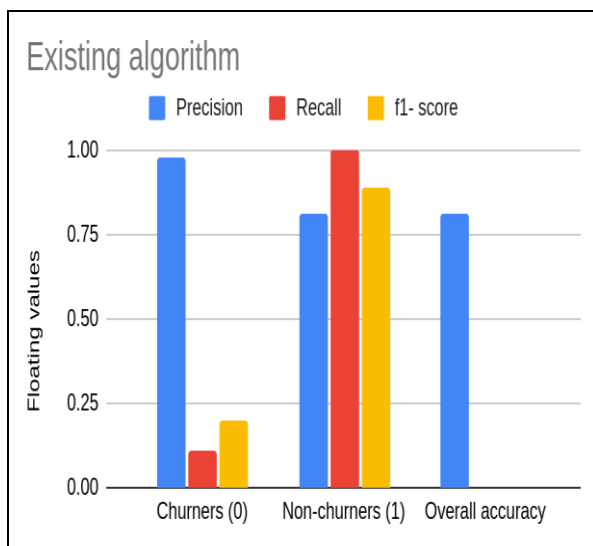
	Existing			Proposed		
	Precision	Recall	f1-score	Precision	Recall	f1-score
Churners (0)	0.98	0.11	0.2	1	0.82	0.9
Non-churners (1)	0.81	1	0.89	0.95	1	0.97
Overall accuracy	0.812			0.95		



(b) Proposed hybrid GWO- KELM algorithm

Figure 3. Comparative graphs between existing and proposed classifiers.

In order to obtain high-quality data, the records are analyzed thoroughly using an EM algorithm. By estimating the probability distribution rate on the set of random variables, it has significantly reduced the redundancies and inconsistencies of the records. It has grouped the policy, policy holder, vehicle, drivers, premiums, claims and claims cost from different sources. In that, missing values are observed on the two important variables, proposal date and date of birth. Based on the computed active years, the authorization details of the drivers are determined. It is resolved by the computed E-step and M-step. In general, the prediction



(a) Existing KELM algorithm

of customer churn turns out to be a complex task, because of the imbalanced data. It is also observed that most of the policy holders stay in a company for a long-term, which also poses another complex task in this study. The role of the EM algorithm concerning to provide balanced datasets are achieved. The ROC analysis presents the achievements of balanced data which is shown in Fig. 2.

Since the analysis of customer behaviors is vague and uncertain, the concept of NB algorithm is applied to find out the similarities of the churners. It is done by building an association between the random set of variables. These values are then imputed in order to find out the similar behaviors of the customers who are willing to become churners. Each behavior of the customers is then fed into the hybrid GWO-KELM algorithm. In Spite of choosing the significant features, the most optimal features are selected by using GWO model. The Fig. 3 & Table 1 presents the achieved results between existing KELM and proposed GWO-KELM. For example, Gender, DOB, sales channel and customer segments are eliminated. Some customers have renewed the policies frequently and thus, duration of the policies are also

discarded. The other features such as Customer ID, motor policies, vehicle type, policy claims, and years insured count and the premium are considered. These selected features have explored what attrition might lead to become a churner. The hybrid algorithm stated that the vehicle type, policy type and the engine size are the attrition of the churners. While, a customer holding more than a policy might also become a churner with respect to the policy renewal. Likewise, the customers who have the highest rate of NCB are less likely to become churners. Compared to the existing KELM classifier, the proposed classifier has obtained 95% prediction accuracy.

5. Conclusion

In this paper, we have introduced a hybrid algorithm, named GWO-KELM that predicts the churns rate in the motor insurance industry using their behavior analysis. Initially, the insurance sector database is acquired and it is preprocessed using Expectation-Maximization algorithm which explores the maximum likelihood rate among the available features. The behavior analysis of their customers are analyzed using Naive Bayes algorithm that associates the relationship between the insurers and the insurance policies. The outcome of the NB

algorithm determines the dependency rate between the churners behavior and the available policies. The feature with dependency rate is fed into the GWO-KELM classifier which extracts the optimal features and finally, classifies the churners and non-churners. Experimental study has suggested efficient predictor variables with the least computational time and efforts.

References

- [1] A. Oyeniyi and A. BAdeyemo, Customer churn analysis in the banking sector using data mining techniques, *African Journal of Computing and ICT*, 8(3), 2015, 166–174.
- [2] N. Kamalraj and A. Malathi, Applying data mining techniques in telecom churn prediction, *International Journal of Advanced Research in Computer Science and Software Engineering*, 310, 2013, 363–370.
- [3] Clara-Cecilie Gunter, IngunnFrideTvette, KjerstiAas, Geir Inge Sandnes, and OrnulfBorgan, Modelling and predicting customer churn from an insurance company, *Scandinavian Actuarial Journal*, 2014(1), 2014, 58–71.
- [4] G. Xia and W. Jin, Model of customer churn prediction on support vector machine, *Systems Engineering-Theory & Practice*, 28(1), 2008, 71-77.
- [5] U. Yabas, H. C. Cankaya and T. Ince, Customer Churn Prediction for Telecom Services, 2012 IEEE 36th Annual Computer Software and Applications Conference, Izmir, Turkey, 2012, 358-359.
- [6] V. Yeshwanth, V. V. Raj and S. Mohan, Evolutionary churn prediction in mobile networks using hybrid learning, Proceedings of the 24th International Florida Artificial Intelligence Research Society, FLAIRS, Florida, USA, 2011.
- [7] M. J. Kim, S. J. Koh and Y. J. Park, A Study on Retaining Existing Customers in the Korean High-Speed Internet Service Market, 2006 Technology Management for the Global Future - PICMET 2006 Conference, Istanbul, Turkey, 2006, 1970-1976.
- [8] J. Cao, H. Zhang and Q. Zheng, Retaining Customers by Data Mining: A Telecommunication Carrier's Case Study in China, 2010 International Conference on E-Business and E-

- Government, Guangzhou, 2010, 3141-3144.
- [9] K. Coussement and D. V. D. Poel, Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques, *Expert Systems with Applications*, 34(1), 2008, 313-327.
- [10] C. Hung and C. F. Tsai, Market segmentation based on hierarchical self-organizing map for markets of multimedia on demand, *Expert Systems with Applications*, 34(1), 2008, 780-787.
- [11] Burez, J., & Van den Poel, D. CRM at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services, *Expert Systems with Applications*, 32(2), 277-288 (2007).
- [12] E. Ascarza, S. A. Neslin, O. Netzer, Z. Anderson, P. S. Fader, S. Gupta, B. G. S. Hardie, A. Lemmens, B. Libai, D. Neal, F. Provost and R. Schrift, In pursuit of enhanced customer retention management: Review, key issues, and future directions, *Customer Needs and Solutions*, 5(1-2), 2018, 65-81.
- [13] D. L. García, À. Nebot and A. Vellido, Intelligent data analysis approaches to churn as a business problem: a survey, *Knowledge and Information Systems*, 51(3), 2017, 719-774.
- [14] B. Larivière and D. V. N. Poel, Predicting customer retention and profitability by using random forests and regression forests techniques, *Expert Systems with Applications*, 29(2), 2005, 472-484.
- [15] L. Guelman, M. Guillén and A. M. Pérez-Marín, Random forests for uplift modeling: an insurance customer retention case, In: Engemann K.J., Gil-Lafuente A.M., Merigó J.M. (eds) *Modeling and Simulation in Engineering, Economics and Management*. MS 2012. Lecture Notes in Business Information Processing, vol 115. Springer, Berlin, Heidelberg, 2012, 123-133.
- [16] J. Vijaya and E. Sivasankar, An efficient system for customer churn prediction through particle swarm optimization based feature selection model with simulated annealing, *Cluster Computing*, 22, 2019, 10757-10768.

- [17] W. Xu, S. Wang, D. Zhang, D and B. Yang, Random rough subspace based neural network ensemble for insurance fraud detection, In Computational Sciences and Optimization (CSO), 2011 Fourth International Joint Conference on Computational Sciences and Optimization, Yunnan, China, 2011, 1276-1280.
- [18] G. G. Sundarkumar, V. Ravi and V. Siddeshwar, One-class support vector machine based undersampling: Application to churn prediction and insurance fraud detection, 2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Madurai, India, 2015, 1-7.
- [19] N. Gordini and V. Veglio, Customers churn prediction and marketing retention strategies. An application of support vector machines based on the AUC parameter-selection technique in B2B e-commerce industry, *Industrial Marketing Management*, 62, 2017, 100-107.
- [20] S. Kim, K. S. Shin and K. Park, An application of support vector machines for customer churn analysis: Credit card case, In: Wang L., Chen K., Ong Y.S. (eds) *Advances in Natural Computation. ICNC 2005. Lecture Notes in Computer Science*, vol 3611. Springer, Berlin, Heidelberg, 2005, 636-647