IJLTET
ONLINE PEER REVIEWED JOURNAL

International Journal of Latest Trends in Engineering and Technology
Vol.(11)Issue(2), pp.029-034
DOI: http://dx.doi.org/10.21172/1.112.06
e-ISSN:2278-621X

PERFORMANCE ANALYSIS OF MACHINE LEARNING ALGORITHMS IN CUSTOMER CHURN PREDICTION

Deepthi Das[1], Dr. Raju Ramakrishna Gondkar[2]

Abstract- Customer attrition is termed by several industrialists and e-commerce professionals to recognize the customers, who are about to change their service from the existing company or end their period of subscription. In recent years, companies such as e-commerce, telecommunication and insurance sectors are facing tremendous pressure due to financial disintermediation and marketing and the gradual increase in the competitiveness tends to provide better service with lesser cost. So, early prediction of the behaviour of the clients plays an important role in the real-time market and can help to retain the loyal customers. In this research, a survey on different data mining techniques and machine learning algorithms along with the challenges of customer attrition prediction in the motor insurance sector are depicted. The survey on the application of the various machine learning algorithm for churn prediction is mainly observed in telecommunication sector and Support Vector Machine (SVM), Artificial Neural Network (ANN) are generally used algorithm for churn analysis and forecasting. Various authors have considered different tools for analysis and the result obtained from the study shows that combination of the two-step process of ANN for training and combined approach of SVM for testing provides better accuracy with high Area Under the Curve compared to existing techniques.

Keywords: Customer attrition, Customer Relationship Management (CRM), Support Vector Machines (SVM), Artificial Neural Network (ANN).

1. INTRODUCTION

Business and corporative enterprises mainly depends on the customers as a key source for income. Although, more importance is given in retaining the existing customers rather than bringing new customers because of several advantages such as easier to explain the product features and selling, building a strong relationship between the company and the existing customers. The customer loyalty is found to be lagging and customer churn is rapidly increasing because of the competitive world, price change and the increasing benefits from the competent company [1]. The Information and Communication Technology (ICT) has developed five strategies for customer life cycle namely acquisition, build up, peak, decline and churn [2]. Customer churn is termed to define a customer who frequently change their supplier in search of new offerings with affordable or less price [3] [4]. Therefore, the process of churn management is a pre-requisite for the growth of the company and helps to access the current situation of the company and setting future plans for the development of the company [5]. The rate of customer churn in an organisation is effectively managed by employing Customer Relationship Management (CRM) persons.

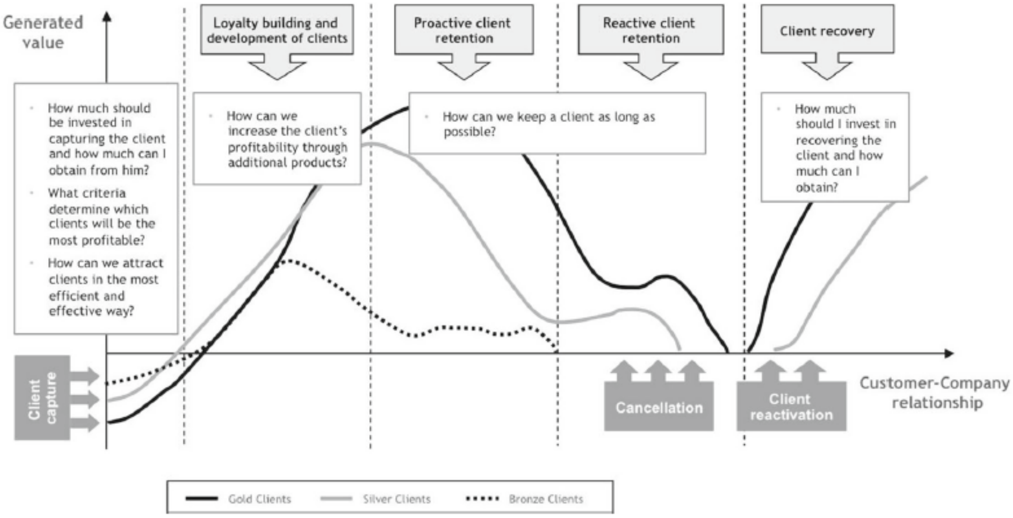
Generated value
Loyalty building and development of clients
Proactive client retention
Reactive client retention
Client recovery
How much should be invested in capturing the client and how much can I obtain from him?
What criteria determine which clients will be the most profitable?
How can we attract clients in the most efficient and effective way?
How can we increase the client's profitability through additional products?
How can we keep a client as long as possible?
How much should I invest in recovering the client and how much can I obtain?
Client capture
Cancellation
Client reactivation
Customer-Company relationship
Gold Clients
Silver Clients
Bronze Clients

Fig. 1. Model deploying relationship with customer and company [9]

[1] Research Scholar, CMR University, Bangalore
[2] Professor, CMR University,Bangalore

Customer Relationship Management (CRM) acts as a bridge between the customer and the organization and the historical behaviour data captured from the customer such as pending bill payment, credit searches issued to evaluate the potential value of the customer churn [6]. The early recognition of customer churn enables the organization to concentrate more on such customers and develop certain specific retention actions to know the reason and retain the profitable and loyal customers. Fig. 1, shows the customer relationship plot of the customer churn frequently observed in the communication and insurance sector and the targeted approaches of customer churn is divided into two types namely, reactive and proactive [7]. During reactive approach, the organization waits until the cancel request for service relationship is obtained from the customer end. An incentive or offer is provided to the customer during their relationship with the organization. During proactive approach, CRM people tries to recognize the customers who are probable to churn before they leave and avoid so by providing extra benefits and incentives to customers. However, the manual prediction is found to be inaccurate and the organization is losing its money by providing benefits to the customers who is not going to churn. Therefore, it is necessary to develop an accurate mathematical model for predicting the rate of customer churn in real time environment [8].

Numerous models comprising number of data mining techniques and machine leaning algorithms have been developed to analyse the behaviour of customer churn in telecom, industrial, and insurance sectors. The churn predictive model plays a vital role during the design of efficient churn management programs and still researches are going to explore more accurate ways of determining the propensities in various real time applications such as telecommunication, insurance, banking and industries. Furthermore, the adoption of digitization technologies leads to rapid increase in the size of the customer input database and in future, an effective and efficient leaning algorithm is required to analyse the big data along with the rate of churn accurately. In this study, an overview of churn prediction models along with machine learning techniques deployed by various authors for early detection of churn rate is presented. The different machine learning algorithms which is frequently used in determining the customer churn is considered for the study and they are Artificial Neural Networks (ANN), Support Vector Machines (SVM), Naïve Bayes Classifier, and Particle Swarm Optimization (PSO).

2. LITERATURE REVIEW

Voluminous amount of research is observed in the field of churn prediction in numerous applications and only few authors has considered their application field as motor insurance since huge churn rate is observed in telecommunication sector. The different authors review on machine learning techniques are as follows

2.1 Random Forest Technique

The prediction of customer churn is found to be a challenging task for Customer Relationship Management (CRM) and the firms is finding difficulties in developing a strong relationship amongst the clients and the industry. The author [10] has developed a churn prediction model using random forests techniques to analyse the behaviour of set of explanatory variables which comprises with details regarding the history of the past customer behaviour, heterogeneity of the customer and several typical constraints of the intermediaries. Customerchurn has been evaluated on the basis of the customer profitability and evolution of the profit. The author considered the real time example of 10,000 customers captured from the input data ware house of European financial service company. The types of random forest technique namely random forest for binary classification and regression forest model for linear dependent variables are employed to analyse the input database. The four type of dependent variables namely next buy, the actively partial defection, the rate of profit drop and evolution of profit is considered for experimental analysis and evaluated in term of Mean Absolute Deviation (MAD). From the study, it is observed that regression forest with MAD value 5.099 for test data set and 4.940 for train dataset outperformed linear regression technique with MAD value of test value of 5.445 and train value of 5.346. Furthermore, the study on research findings show that regression technique provides better fitness value for the estimation and analysis of input data and the similar set of variables have distinct impact amongst features such as buying vs regression vs behaviour of profit. The author summarizes stating that historical data of the customer has greater impact in determining the behaviour of the customer and can be used to predict the churn rate of the customer. The advancement of random forest is used to analyse the problem of uplift modelling and to calculate the probability of customer switching to other company [11]. The uplift model is characterized through the number of input vectors or the predictive variables and processed to identify the behaviour of the customer and retain the valuable customer. The prediction of uplift is captured through the average uplift value from each tree in the ensemble and it be varied by using two parameters namely the presence of set of variables in the random subset at individual node a units of tress in the forest. From the study on experimental analysis, it is observed that the retention rate is calculated through the model inputs and the result depict that an overall retention rate of 91.3% is achieved by considering several vehicle characteristics such as type of the vehicle, age, price and lease of the vehicle. Several driver characteristics such as accidents due to the fault, non-fault and driver age and by considering policy characteristics such as state of premium, endorsements and coverage. Furthermore, it is observed that the probability of customer churn strongly depends on the age and the history of the claim. The growth in the market development has given more importance to churn since the assets of the customer are the key source of company growth and income. The author [12] has developed a hybrid model comprising random forest and support vector machine to predict the behaviour of churn in a organization.

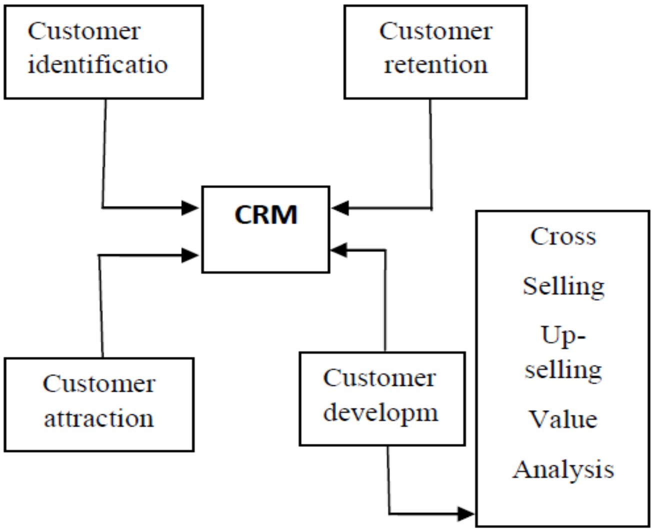

Figure. 2. External factor considered by CRM [12]

Figure2, shows the factors considered by author for the prediction of rate of churn. Thorough monitoring and evaluation of the above mentioned variables can increase the performance of the system. The author has considered 3333 instances and a set of 21 attributes to evaluate the performance of the system and it is implemented using MATLAB u2013a tool. The results obtained from the study shows that an accuracy of 97.19% is achieved from the hybrid model with an error rate 2.50% and comparative analysis shows that developed method outperformed other existing techniques.

2.2 Particle Swarm Optimization

The Particle Swarm Optimization (PSO) is found to be advantageous in solving the recursive problems and provides solutions to the difficulty in prediction through a set of applicant solutions namely particles. An effective churn predictive model comprising large set of telecom data is analysed using a metaheuristic based churn prediction technique [13]. The rate of churn is found to be extreme in communication sector and the author has considered PSO since it uses metaheuristic technique to obtain the set of fitness parameter solution through the selection of candidate solutions namely particles. The author has implemented PSO through C#.Net on Visual Studio 2012 platform. The orange benchmark dataset from a French telecom company is considered as an input database and processed through PSO for predictive analysis. From the study, it is observed that the developed PSO model is analysed in terms of true positive and true negative rate and the results through the Region of Convergence curve showcase that the developed model provides best accuracy rate compared to other existing techniques. Furthermore, an enhanced prediction model to calculate the rate of churn in telecom industry is developed by author [14]. CRM is found to be a major aspect of the organization and a prediction of churn has gained key prominence in recent times due to the increase in competitors and the decrease in price with additional benefits. The author has considered orange dataset with 230 attribute density and 50,000 instances as input database andPSO algorithm for processing and calculation of prediction rate. The results obtained from the study on experimental evaluation shows that PSO exhibits an accuracy rate of approximately 85% with an F_Measure rate of approximately 0.5, which is better compared to existing techniques with good positive retrieval rates. Furthermore, an effective and efficient customer churn predictive system comprising Particle Swarm Optimization (PSO) on the basis of feature selection along with simulated annealing is proposed by author [15]. The manual analysis of customer churn is found to be difficult due to the globalization of market and PSO is found to be a better alternative solution. The author has considered several factors such as volume of the data, data imbalance, selection of key attributes and highly sparse data to analyse the behaviour of the system. Furthermore, combination of several models such as PSO and feature selection technique, PSO and simulated annealing, PSO along with feature selection and simulated annealing is considered for the analysis and performance evaluation is achieved through the parameters such as accuracy, true positive rate, true negative rate and precision. From the study, it is observed that the performance is evaluated in terms of 1000 samples, 5000 samples and 7344 samples and the results obtained shows that high accuracy rate of 90.65% is obtained through PSO-FSSA with an true positive rate of 92.92% and true negative rate of 77.41% which is better and efficient compared to other existing techniques.

2.3 Support Vector Machines

The early warnings of customer churn along with fraud detection and developing an effective model is found to reduce the financial losses of approximately millions of in several sectors [16]. The advantages of SVM over other techniques are minimal input parameters for optimization, linearity constrained quadratic problem for obtaining training dataset and development of technique is mainly dependent on structural risk minimization [21]. The under sampling approach using one class Support Vector Machine (SVM) is discussed to ascertain the churn prediction in credit card and automobile insurance sector [17]. The study shows that two set of data with a ratio of 80.0% and 20.0% is extracted which comprises with a details of 738 fraudulent in insurance sector. Further, the data is classified and processed through one class SVM for pre-processing and classification. From the study, it is observed that an accuracy rate of 60.61% is achieved with higher Area Under the Curve (AUC) value of 8728.5. Furthermore, the author compared the developed technique with the model proposed by author [18] and the study on comparative analysis shows that developed under sampling methodology increases the rate of AUC and reduces the complexity rate compared to existing techniques. Furthermore, the application of SVM is used in the field of B2B e-commerce industry to predict the rate of churn and through unique testing and forecasting models [19]. The advancement in the e-commerce sector has provided a distinct solution for sharing information between the clients and the business. The author has captured input dataset from the online marketing database website [20]. The capability of RBF kernel technique to handle relationship amongst the dependent and independent variable set is used to optimize the layer of hyper plane and the data processed is forwarded through SVM model for classification. MATLAB software is used for analysis and the result obtained from the study shows that SVMauc has highest prediction rate of 89.98% with AUC of 88.61, which is better compared to other existing techniques. From the summarization section, it is observed that the predictive performance can be further improvised by comprising the staying power of the system. Furthermore, a predictive model for customer churn by using the selective item sets such as bill, records of the data and terminals and their features sets is developed by author [22]. From the survey conducted by author, it is shown that SVM is found to be better applicable to analyse customer churn analysis and results obtained from the study shows a reasonable accuracy rate of 70.6% is obtained for 403280 input data.

2.4 Artificial Neural Network

The churn prediction and management is found to be a major issue and the author [23] has developed a hybrid model comprising the combination of ANN and ANN (Artificial Neural Network) for pre-processing and combination of Self Organised Maps (SOM) and ANN for predictive analysis. The single layer and multilayer perceptron is found to be effective during the training set and the problems of clustering can be easily analysed through this process. The author has considered two stage process of ANN to achieve higher rate of accuracy during the training phase and SOM and ANN for analysing clustering and to achieve higher prediction accuracy. The result obtained from the study shows that an accuracy rate of 93.06% is achieved for testing model and 94.32% during training. In general, it is observed that the developed model successfully predicts the type-1 and type-2 error and better results is observed at the training stage rather than testing stage. Furthermore, the combination of K-means technique for clustering and Multi-Layer Perceptron based Artificial Neural Network (MLP-ANN) for prediction is developed by author [24]. The author has developed three models namely k-means, hierarchical and SOM on the basis of real time churn data and processed through ANN for classification purpose. The result obtained from the study shows that K-means along with MLP based ANN predicts better accuracy rate of 97.2% with churn rate of 73%. Moreover, the slightly higher churn rate of 94.6%is observed by deploying combination of SOM and MLP-ANN model, but the accuracy rate is decreased to 95.9%. Further, the author [25] has considered the analytical behaviour of the customer to predict the rate of churn using machine learning techniques. Due to the advancement and modification in the approaches of selling and supply surplus has created a huge amount of competition in the market. Six stages namely retrieval of customer data and formation of dataset, pre-processing and selection of key characters, differentiating into training and testing database, usage of effective machine leaning techniques for prediction model development, comparative analysis followed by deploying appropriate strategies for performance evaluation is used in the analysis. The result obtained from the study shows that ANN-MLP model provides highest accuracy rate of 97.07% and F_measure rate of 97.92% compared to ANN-RBF and SVM-RBF.

3. COMPARATIVE ANALYSIS

A comparative table is obtained by using the advantages of the machine learning techniques deployed by several authors as shown in the aforementioned research and compared on the basis of performance parameters which are depicted as shown in Table 1.

Table 1. Comparative analysis of machine learning techniques for customer attrition prediction

Author	Classifier	Accuracy Rate	Comments
Mahajan D et al., [12]	Combined approach of random forest and support vector machines	97.19%	The result obtained is for 3333 input variables with a set of 21 attributes. Further, less error rate of 2.50% is observed through the analysis
Vijaya, J et al., [15]	Combined approach of Particle swarm optimization and Feature selection simulated annealing	90.65%	The developed model is tested for 7344 samples and the results obtained shows that there is an increase in precision with true positive rate of 92.92% with true

			negative rate of 77.41%.
Coussement, K., et al., [20]	Support Vector Machines	89.98%	High prediction rate is achieved with AUC rate of 88.61
Khodabandehlou, S., et al., [25]	Multi-Layer Perceptron based Artificial Neural Network	97.07%	F_measure rate of 97.92% is achieved showing higher true positive rate and predictiveness.

Figure 3. Performance analysis

From Figure 3, it is observed that combined approach of RF and SVM and another combined approach model of MLP and ANN predicts better accuracy rate. The author [23] stated that the two-step process using hybrid ANN model provides a better feature set for churn rate detection. The analysis obtained through the combination of RF along with SVM also provides better accuracy rate. In future, better attrition prediction model can be developed by using the combinational approach of ANN and SVM for the application of motor insurance sector.

4. CONCLUSION

The primary aim of this research findings is to detect an efficient machine learning technique to classify the customer attrition on the basis of the number of regular and non-regular customers of an insurance company. The aforementioned research ascertains different authors review on input data selection, pre-processing steps, selection of algorithms and their performance evaluation in terms of accuracy. The analysis on review shows that many machine learning techniques are developed in terms of churn prediction for telecommunication sectors and minimal amount for motor insurance section. Through the results obtained from the review, it is observed that ANN and SVM have a better probability of churn prediction and it can be used for the application motor insurance sector.

5. REFERENCES

[1] Xia, G. E., &Jin, W. D. Model of customer churn prediction on support vector machine. Systems Engineering-Theory & Practice, 28(1), 71-77 (2008).

[2] Yabas, U., Cankaya, H. C., &Ince, T. Customer Churn Prediction for Telecom Services. In Computer Software and Applications Conference (COMPSAC), 2012 IEEE 36th Annual. pp. 358-359. IEEE (2012).

[3] Yeshwanth, V., Raj, V. V., &Saravanan, M. Evolutionary churn prediction in mobile networks using hybrid learning. In Twenty-Fourth International FLAIRS Conference (2011).

[4] Kim, M. J., Koh, S. J., & Park, Y. J. A Study on Retaining Existing Customers in the Korean High-Speed Internet Service Market. In Technology Management for the Global Future, 2006. PICMET 2006. Vol. 4, pp. 1970-1976. IEEE (2006).

[5] Cao, J., Zhang, H., & Zheng, Q. Retaining Customers by Data Mining: A Telecomunication Carrier's Case Study in China. In E-Business and E-Government (ICEE), 2010 International Conference on (pp. 3141-3144). IEEE (2010).

[6] Hung, C., & Tsai, C. F. Market segmentation based on hierarchical self-organizing map for markets of multimedia on demand. Expert systems with applications, 34(1), 780-787 (2008).

[7] 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).

[8] Ascarza, E., Neslin, S. A., Netzer, O., Anderson, Z., Fader, P. S., Gupta, S., ...& Provost, F. In pursuit of enhanced customer retention management: Review, key issues, and future directions. Customer Needs and Solutions, 5(1-2), 65-81 (2018).

[9] García, D. L., Nebot, À.,&Vellido, A. Intelligent data analysis approaches to churn as a business problem: a survey. Knowledge and Information Systems, 51(3), 719-774 (2017).

[10] Larivière, B., & Van den Poel, D. Predicting customer retention and profitability by using random forests and regression forests techniques. Expert Systems with Applications, 29(2), 472-484 (2005).

[11] Guelman, L., Guillén, M., & Pérez-Marín, A. M. Random forests for uplift modeling: an insurance customer retention case. In Modeling and Simulation in Engineering, Economics and Management pp. 123-133. Springer, Berlin, Heidelberg (2012).

[12] Mahajan, D., &Gangwar, R. IMPROVED CUSTOMER CHURN BEHAVIOUR BY USING SVM (2017).

[13] T. Sumathi. Churn Prediction on Huge Sparse Telecom Data Using Metaheuristic.International Journal of Advanced Research in Computer and Communication Engineering, pp. 574-577 (2016).

[14] R.Suganji, Dr. G. Ravi. Enhanced Churn Prediction on Huge Telecom Data using Particle Swarm Optimization. International Journal of Advanced Research in Computer and Communication Engineering pp. 2673-2677 (2016).

[15] Vijaya, J., &Sivasankar, E. An efficient system for customer churn prediction through particle swarm optimization based feature selection model with simulated annealing. Cluster Computing, 1-12. (2017).

[16] Xu, W., Wang, S., Zhang, D., & Yang, B. Random rough subspace based neural network ensemble for insurance fraud detection. In Computational Sciences and Optimization (CSO), Fourth International Joint Conference on pp. 1276-1280. IEEE (2011).

[17] Sundarkumar, G. G., Ravi, V., &Siddeshwar, V. One-class support vector machine based undersampling: Application to churn prediction and insurance fraud detection. In Computational Intelligence and Computing Research (ICCIC), 2015 IEEE International Conference on(pp. 1-7). IEEE (2015).

[18] Sundarkumar, G. G., & Ravi, V. A novel hybrid undersampling method for mining unbalanced datasets in banking and insurance. Engineering Applications of Artificial Intelligence, 37, 368-377 (2015).

[19] Gordini, N., &Veglio, V. 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, 100-107 (2017).

[20] Coussement, K., & Van den Poel, D. Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. Expert systems with applications, 34(1), 313-327 (2008).

[21] Kim, S., Shin, K. S., & Park, K. An application of support vector machines for customer churn analysis: Credit card case. In International Conference on Natural Computation (pp. 636-647). Springer, Berlin, Heidelberg (2005).

[22] Dong, R., Su, F., Yang, S., Cheng, X., & Chen, W. Customer Churn Analysis for Telecom Operators Based on SVM. In International Conference on Signal And Information Processing, Networking And Computers (pp. 327-333). Springer, Singapore (2017).

[23] Tsai, C. F., & Lu, Y. H. Customer churn prediction by hybrid neural networks. Expert Systems with Applications, 36(10), 12547-12553 (2009).

[24] Hudaib, A., Dannoun, R., Harfoushi, O., Obiedat, R., &Faris, H. Hybrid data mining models for predicting customer churn. International Journal of Communications, Network and System Sciences, 8(05), 91 (2015).

[25] Khodabandehlou, S., and Zivari Rahman, M. Comparison of supervised machine learning techniques for customer churn prediction based on analysis of customer behavior. Journal of Systems and Information Technology, 19(1/2), 65-93 (2017).