WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Customer Churn Prediction System

B.V.Niranjani[1], Mr.R.Sathishkumar.,MCA.,M.Phil.[2]

[1]Master of Computer Application, krishnasamy college of engineering &Technology,Cuddalore

[2]Assitant Professor, Master of Computer Application, krishnasamy college of engineering &Technology,Cuddalore

ABSTRACT

Customer churn is an extensively used term in several organizations such as gaming, telecommunications and banking. Churn denotes the act of a user leaves a product or services. To avoid churn, the customer activities will be followed, recognizing the potential interest to the customer and employ those data to adopt push messages personally in order to draw an attention to customer back to the product or services. In recent days, the Customer Churn Prediction (CCP) becomes more popular and several techniques have been proposed for effective prediction. The CCP can be considered as a data classification problem, which efficiently classifies the presence of churn, or not. Presently, different CCP models has been developed using data mining, artificial intelligence techniques, and so on. In this paper, we attempted to analyse the performance of different data mining techniques such as Support Vector Machine (SVM), Random Forest Classifiers on the CCP. Using a benchmark dataset, the applied algorithms are validating in terms of different performance measures. The experimental results depicted that the Random Forest Classifier is found to be effective than the other classifiers.

INTRODUCTION

In an era of mature markets and intense competitive pressure, it is fundamental for companies to manage relationships with their customers, in order to increase their revenues. In business economics, this concept known as "Customer Relationship Management" (CRM), which is a business strategy that aims at ensuring customers' satisfaction. The companies that successfully apply CRM to their business nearly always improve their retention power, that is, the probability that a customer will not leave. In fact, a high retention power avoids a useless waste of money since acquiring new customers can cost many times more than satisfying and retaining existing customers. This prompted businesses to invest heavily in behaviour analysis in order to predict which clients would leave ahead of time.

In particular, these customers are subjected to several rectifying actions (e.g. promotions or gifts) for their retention.

The phenomenon related to the customer's abandonment is commonlyas "churn," while the churn detection process is usually named as "Customer Churn Prediction." In this thesis, we focus on the rest phase only that is, the correct identification of the customers who will leave the company (namely the "churners").

Technically speaking, we chose to model the churn prediction problem as a standard binary classification task, labelling each customer as either "churner" or "non-churner."

An important byproduct of our study is that we formally identify who is a "churner" according to the bank's needs. Several definitions can be applied to label a customer as a "churner," such as a customer who closes all their bank accounts or who has a very low balance and does not make transactions for a long time. In particular, we labelled as a churner in a given calendar month a customer who leaves during the subsequent month, to predict sufficiently in advance which customers will leave. In this way, banks have the chance to take some corrective actions useful for retaining their customers.

The obtained results confirm that machine learning techniques can be effectively applied to predict customer churn in banking environments, that is, succeeding in forecasting which customers will leave in the near future with high precision (i.e., avoiding false positive costs). Hence, more in-depth studies may be worth pursuing.

Presently, different CCP models have been developed using data mining, artificial intelligence techniques, and so on. The non-existence of review papers in the field of CCP inspired us to carry out the study. In this paper, a review of different CCP models has been conducted based on objectives, techniques used, merits, performance, and so on. A detailed comparison of the reviewed techniques is also made, and the reviewed methods are briefly discussed.

EXISTING SYSTEM

Various types of optimization techniques are applied for different types of segmentation, such as Particle Swarm Optimization (PSO), Optics Inspired Optimization (OIO), Biogeography Based Optimization (BBO) and Genetic Algorithm (GA). Before being applied to a specific problem, all evolutionary and swarm intelligence-based algorithms require parameter descriptions, including population size, the dimension of individual population members, and predefined algorithm-dependent parameters. The performance of an algorithm for capturing approximate solutions depends on the fine-tuning of algorithm parameters. Rashedi et al. proposed a Gravitational Search Algorithm (GSA) inspired by the law of gravity. It was observed that GSA performs better than well-known optimization techniques such as PSO, GA, and SA when it is tested on various benchmark functions. This is the motivation to apply GSA to image segmentation in the proposed work. The flow of GSAis presented and can be described as follows: the law of gravity inspires this algorithm. The search agents are modelled as a collection of objects, which interact with each other based on Newtonian physics. Every mass represents a solution and the algorithm has to adjust between gravitational and inertial mass. The heaviest of them all will attract the masses, which will present an optimum solution in search space. The force acting on the heaviest object drifts it apart from the rest of the population, which is the optimal solution.

It consists of six phases, namely

I. Identification of most suitable data(variance analysis, correlation matrix, outliers removal, etc.)
II. Cleaning & Filtering (handling null and missing values)
III. Feature Selection(Gravitational Search Algorithm)
IV. Development of predictive models(Logistic Regression, Support Vector Machine, Naïve Bayes)
V. Cross validation(k-fold Cross validation)
VI. Evaluation of results(Confusion matrix, AUC curve)

We have applied gravitational search algorithms to perform feature selection and to reduce the dimensions of the dataset in existing approaches where prediction accuracy is low due to improper feature selection. The ADABOOST and XGBOOST models have limitations on over fitting problems.

PROPOSED SYSTEM

WOA is a metaheuristics algorithm. As implied by the name, this algorithm was proposed for optimising numerical problems by Mirjalili and Lewis (2016). The inspiration for WOA comes, in particular, from social behaviour and the bubble-net hunting of humpback whales in the oceans.

The process of this proposed system includes four steps they are.

First, Load Data is the data that we need to load to start any of the ML projects. Concerning data, the most common format of data for ML projects is CSV (Comma-Separated Values). CSV is a simple file format which is used to store tabular data (number and text), such as a spreadsheet in plain text.

Second, preprocessing is a transformation applied to our data before feeding it to the algorithm. Data preprocessing is a technique that is used to convert raw data into a clean data set.

Third, WOA Feature Selection is a Novel Whale Optimization Approach for Feature Selection. Crossover and mutationare used in exploitation properties in the WOA algorithm. Tournament selection is used in the WOA algorithm to improve exploration. The experiments prove the superior performance of the proposed approaches.

Fourth, PSO Random Forest in this PSO is a computational method that optimises a problem by iteratively trying to improve a candidate solution about a given measure of quality.

Random Forest is a supervised machine learning algorithm made up of decision trees. A random forest is used for both classification and regression.

At present, the business services provided by major telecommunications companies to their customers are similar, with no distinctive features. With the development of society and the increasingly fierce competition in the industry, customers' demand for personalised services is getting higher and higher, resulting in frequent customer churn. Telecommunications urgently need to adjust their business strategies and develop programmes to meet customers' needs for personalised services so as not to lose customers and retain economic benefits. Customer churn prediction analysis is an important part of customer relationship management.

The paper uses an improved whale optimization algorithm to improve the support vector machine to get an improved WOA-SVM model. It is also compared with GA-SVM, SVM, multilayer perceptron, and logistic regression methods with an example of VIP customer churn prediction in domestic commercial telecommunication, and it is found that the model obtains the best correct rate, hit rate, coverage rate, and boost factor, and is an effective model for predicting customer churn in telecommunications.

VISUALIZATION

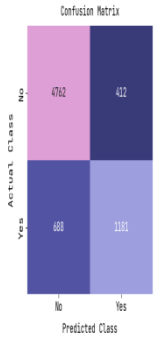

Figure1: CCP Prediction Result

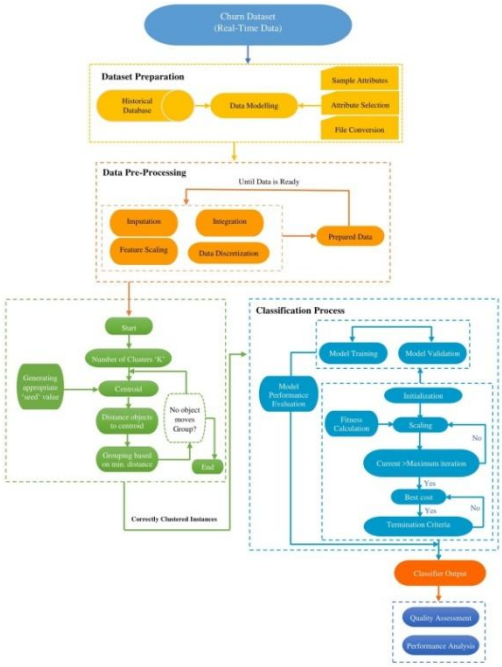

Figure2: Architecture

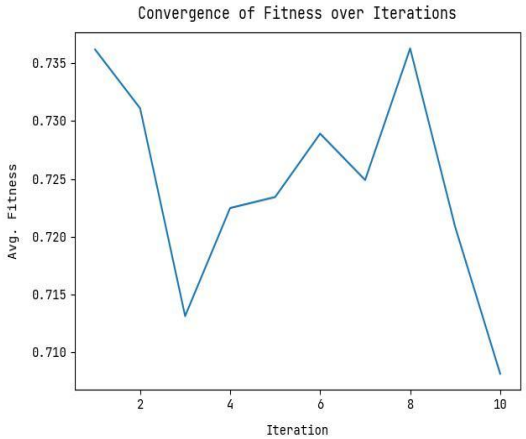

Figure3: Convergence Graph of WOA

DATASET

It contains information about the customer's services and contact information such as customer id, gender, senior citizen, partner, dependents, tenure, phone service, multiple lines, internet service, online security, online backup, device protection, tech support, streaming TV, streaming movies, contract, paperless billing, payment method, monthly charges, total charges, churn. This data has a large size and there is a lot of detailed information about it. We spent a lot of time trying to understand it and to know its sources and storing format. In addition to these records, the data must be linked to the detailed data stored in relational databases that contain detailed information about the customer.

CONCLUSION

In recent days, customer churn prediction (CCP) has become popular, with several techniques having been proposed for effective prediction. The CCP can be considered as a data classification problem, which efficiently classifies the presence of churn, or not. In this paper, we reviewed several CCP models developed for different fields concerning their objectives, methodology used, benefits, and performance. The CCP methods developed for different domains such as TCI, the financial sector, online gaming, social media, etc. are discussed. A detailed comparison of reviewed approaches is also made based on the objectives, methodology used, application area, measures used, and experimental results. In the telecommunication business, Customer Churn Prediction (CCP) is a significant problem.

In this business circumstance, a trustworthy user predictor will be considered a valuable one. Keeping this in mind, in this paper, we attempted to analyse the performance of different data mining techniques like SVM, and Random Forest on the CCP, Using a benchmark dataset, the applied algorithms are validated in terms of different performance measures. The experimental results showed that the Random Forest Classifier was found to be more effective than the other classifiers.

REFERENCES

[1] Coussement, K., Lessmann, S., &Verstraeten, G. (2017).A comparative analysis of data preparation algorithms for customer churn prediction: A case study in thetelecommunication industry. Decision Support Systems, 95, 27–36. http://dx.doi.org/10.1016/j.dss.2016.11.007.

[2] Óskarsdóttir, M., Bravo, C., Verbeke, W., Sarraute, C., Baesens, B., &Vanthienen, J.(2017). Social network analytics for churn prediction in telco: Model building,evaluation and network architecture. Expert Systems with Applications, 85, 204–220. http://dx.doi.org/10.1016/j.eswa.2017.05.028.

[3] Sharma, R. R., &Rajan, S. (2017). Evaluating prediction of customer churn behaviour based on artificial bee colony algorithm. International Journal Of Engineering AndComputer Science, 6(1), 20017–20021. http://ijecs.in/issue/v6-i1/32ijecs.pdfhttp://dx.doi.org/10.18535/ijecs/v6i1.32.

[4] Farquad, M., Ravi, V., &Raju, S. B. (2014). Churn prediction using comprehensiblesupport vector machine: An analytical CRM application. Applied Soft Computing, 19,31–40.

[5] Burges, C. J., Efimov, D., & Darrow, R. (1998).Airline new customer tier levelforecasting for real-time resource allocation of a miles program. Data Mining andKnowledge Discovery, 2(2), 121–167. http://link.springer.com/10.1023/A:1009715923555http://dx.doi.org/10.1023/A:1009715923555.

[6] Ganesh, J., Arnold, M. J., & Reynolds, K. E. (2000). Understanding the customer baseof service providers: An examination of the differences between switchers andstayers. Journal of Marketing, 64(3), 65–87

[7] Garcia, D. L., Angela, N., &Vellido, A. (2017). Intelligent data analysis approachesto churn as a business problem: A survey. Knowledge and Information Systems,51(3), 719–774. http://dx.doi.org/10.1007/s10115-016-0995-z.

[8] Amin, A., Anwar, S., Adnan, A., Nawaz, M., Howard, N., Qadir, J., ...Hussain, A.(2016). Comparing Oversampling Techniques to Handle the Class ImbalanceProblem: A Customer Churn Prediction Case Study. IEEE Access,4.
http://dx.doi.org/10.1109/ACCESS.2016.2619719

[9] Athanassopoulos, A. (2000). Customer satisfaction cues to support marketsegmentation and explain switching behavior. Journal of Business Research, 47(3),191–207

[10] Maria, O., Verbeke, W., Sarraute, C., Baesens, B., &Vanthienen, J. (2016).Acomparative study of social network classifiers for predicting churn in thetelecommunication industry. IEEE/ACM International Conference on Advances inSocial Networks Analysis and Mining (ASONAM) (pp. 1151–1158).

[11] Jamil, S., & Khan, A. (2016). Churn comprehension analysis for telecommunicationindustry using ALBA. ICET 2016–2016 International Conference on EmergingTechnologies (pp. 1–5). . http://dx.doi.org/10.1109/ICET.2016.7813259.

[12] Idris, A., Khan, A., &Soo, Y. (2013). Intelligent churn prediction in telecom:Employing mRMR feature selection and RotBoost based ensemble classification.Springer Science Business Media New York (pp. 659–672).

[13] Amin A, Al-Obeidat F, Shah B, Adnan A, Loo J, Anwar S. Customer churn prediction

in telecommunication industry using data certainty. Journal of Business Research.2018 Mar 19.

[14] Stripling E, vandenBroucke S, Antonio K, Baesens B, Snoeck M. Profit maximizinglogistic model for customer churn prediction using genetic algorithms. Swarm andEvolutionary Computation. 2018 Jun 1;40:116-30.