

A Hybrid Classification Approach for Customer Churn Prediction using Supervised Learning Methods: Banking Sector

Putta Hemalatha, Geetha Mary Amalanathan*

School of Computer Science and Engineering

Vellore Institute of Technology, Vellore, Tamil Nadu, India

putta.hemalatha2016@vitstudent.ac.in ; geethamary@vit.ac.in

Abstract—The substantial efforts kept by many companies in retaining customers has become more challengeable. In this aspect, forecasting a customer to churn in the upcoming future is an enormously prevailing job for many advertising crews. Present market suffers much-sophisticated charges in bidding to handle new-fangled customers than to hold prevailing ones. As a consequence, ample exploration has been capitalized to innovative traditions of recognizing individual customers who may consume a high hazard of agitating. Nevertheless, purchaser preservation pains consume huge volumes of an organizational resource. In accordance with these matters, the succeeding generation of churn administration ought to concentrate on down truth. A diverse technique of churn prediction has been established to fulfill the above requirements. The work that carried out in this paper is to develop a model that can predict better accuracy of churn. We used the existing models of SVM and kernel function of it and collaborated with the artificial neural networks in prediction. The results proved to be outstanding when both the models integrated with each other. Thus, most of the widespread innovations that have been distinguished in the writing for the extension of a customer churn administration stages which are using the soft computing techniques such as Neural networks, SVM etc.

Keywords: *churn prediction, customer retention, neural networks, support vector machine.*

I. INTRODUCTION

The latest development in the banking sector made the machine learning to cope up with the business needs. Various business improvements turned out to be creative and project a plan of action in expanding the cost of customer securing. Churn prediction stands have an important consideration in customer relationship management (CRM). In the current competitive business environment, customers tend to switch between companies easily. Researchers in their study have observed that acquiring a new customer is generally six times more expensive than retaining an existing customer [1].

The suppliers are required to put additional endeavors in forecasting and counteractive action to identify churners. This paper means to exhibit normally utilized data mining strategies for the attribute churn. In view of the current bank dataset, the proposed techniques attempt to discover designs which can bring up conceivable churners. Many organizations such as telecommunications, financial services, airline ticking services, banking sector, and online gaming are much

keen in developing a strong customer relationship with their customers which in turn bring fruitful profits to the companies [2].

A review paper named, Intelligent Data Analysis Approaches to Churn as Business Problem by researcher David L. Garcia et.al [7] addressed that customer retention leverages in understanding customer loyalty and further build's mechanisms in order to anticipate customers' intention to churn.

The mining techniques applied in CRM is now an emerging trend which gained much attention in the researchers to investigate the customer behavior and respected aspects to improvise the business strategies [8]. The supervised algorithms utilized for forecasting are Neural Networks, SVM and KSVM. Customer churn forecast has been constrained to demonstrating churn in the following (attributes) day and age. We recommend an active system for forecasting churners by utilizing information from a customer's database, and use it for foreseeing churned customer by exploiting standard classifiers [3].

The proposed system which is a combination of KSVM and ANN is performing better when compared to single model predictions, which incorporates customer perceptions from various generations, and consequently tends to the outright irregularity issue that is important for the most profitable customer portion of many organizations.

In the proposed paper the researcher has built a hybrid (KSVM + ANN) approach to discover churn customers in private sector banks and the evaluated result will compare with three state-of-the-art algorithm SVM, KSVM, and ANN. The research was carried out on a banking customer database reserved by random sampling. Section 2 presents reviews of literature, followed by system adopted models, section 3 comprises of system architecture, in section 4 methodology of the paper is observed, section 5 deals with proposed work, section 6 comprises of results and discussion, finally section 7 ends with conclusion and future work and at last references are given by the researcher.

II. LITERATURE REVIEW

Many researchers have come up with different approaches such as hybridizing machine learning algorithms such as SVM, Naive Bayes, Neural networks etc.[4].

Nijhawan et.al [2] have demonstrated in the present setting where individuals are given an enormous selection of offers and distinctive specialist co-ops to settle on, winning new customers is an exorbitant and hard process. Along these lines, putting more exertion in keeping churn low has turned out to be basic for banks and administration situated organizations.

Regularly, CRM applications hold an enormous arrangement of data with respect to every individual customer. This data is picked up from customers' action at the organization, information entered by the customer during the time spent enlistment, calls to help hotlines, and so forth. Legitimate examination of this information can bring wonderful outcomes for advertising purposes, additionally to identify customers who are probably going to cross out their agreement [10]. CRM plays a key role in business development. There are different dimensions in which we can relate it, in understanding the overall performance development of a customer [7].

A. Customer Retention

Customer Retention is quite critical to maintaining a vigorous on-going association with consumers. In this aspect, the objective should change to emerging a customer-centric background, which instructs perceptions that provide better services and fulfill the requirements of a customer [8]. Consequently, the corporate leaders have to begin thinking more on the thought-processes of dealers to enhance the income scale. The objective is to expand customer procurement requiring little or no effort, without endangering existing customer relations and guaranteeing the smooth move of another customer to a devoted.

B. Machine Learning (ML) Techniques for Customer Churn Prediction

There are various ML techniques which have proven to produce better results in customer churn prediction which are discussed below in brief

Classification is one of the data mining methods which depend on machine learning systems to arrange things in an informational index into a predefined set of classes or gatherings [9]. The process of classification adopts techniques of mathematical and statistical ways which are applied to decision trees, linear programming, neural networks and support vector machines etc.

The mentioned ways are discussed below in short.

1) ANN's In Modeling Customer Churn Prediction

ANNs especially MLP network as a suitable model specifically, statistic, utilization and charging information is the greatest helpful decision for creating a customer agitate forecast philosophy for the accompanying factors[10]. Study of a customer's standard use conduct and checking action, sudden diminished use can be deciphered as a customer's objective to churn.

A multilayer perceptron (MLP) is a special class

of feedforward artificial neural network. An MLP consists of; three layers, namely an input layer, a hidden layer, and an output layer. This is a supervised learning technique called backpropagation for training. The distinguished data is not linearly separable.

There are many thumb rule methods in identifying the required number of neurons to set in the hidden layers; here we used the following measure to calculate it

The hidden neuron size

$$N_h < \frac{2}{3} N_i + N_o \quad (1)$$

Where, N_h = Hidden neuron size

N_i = Input layer size

N_o = Output layer size

2) SVM Modeling For Churn Prediction

Chandrakala [4] expressed that there is a significant group of work on agitate expectation models. The underlying concentration of the writing was an examination of characterization strategies as far as expectation precision procedures, for example, Decision trees, Logistic regression, neural systems, support vector machines, and survival regression are the most mainstream techniques. While decision trees and SVM have been utilized for prescient purposes, survival regression is utilized and had done on the later applications utilizing half and half models or gatherings that coordinate numerous classifiers as well as create variations of the current calculations which enhance forecast exactness while diminishing interpretability. In their vast logistic regression De Caigny et.al [5] execution of numerous classifiers is tantamount since various classifiers produce improved exhibitions in various settings and data records.

3) K-SVM Model

Kernel function of SVM is the part show of support vector machine and it is a standout amongst the most generally utilized characterization techniques [6]; the use of kernel with regularization gives better performance. The radial function is given by

$$k(x, \acute{x}) = \exp\left(\frac{\|x - \acute{x}\|^2}{2\sigma^2}\right) \quad (2)$$

Where x, \acute{x} are two samples in the featured vector space, σ is a free parameter, $\|x - \acute{x}\|^2$ represents Euclidean distance.

The above-mentioned equations are used according to the requirement of the research.

III. SYSTEM ARCHITECTURE

The system architecture is an abstract model that defines how the work is carried out in each and every phase of the system. The first phase of the system explores the identification and collection of the database required to carry out the work. In the second phase, the data semantics are developed and are feed as input to the third phase of the system. Accordingly, the third phase selects the features that are best suitable for predicting the model requirements followed by the fourth phase which

sets the stage for strong development of the model for accurate prediction. Finally, the last phase validates the results and forwards the results to the selection process to fetch again better features for prediction. Thus, the system is proven to be successful in every phase in fulfilling the objectives of the work accomplished.

IV. METHODOLOGY

The outcomes from this examination give an establishment in writing review successfully. Fig.1 visualizes the overall work carried out in a data flow diagram representation. Goals are distinguished through the review of writing. With the point and targets characterized, advancement of philosophy for churn expectation is conceivable. Once a system has been made a product model is developed. Approval of the approach is performed utilizing the product model. The outcomes acquired through approval are contrasted and two prevalent agitate forecast systems distinguished through the writing study. Impediments of the examination are closed through approval and agitate expectation procedure correlation. This exploration has along these lines concentrated on quantitative research strategies.

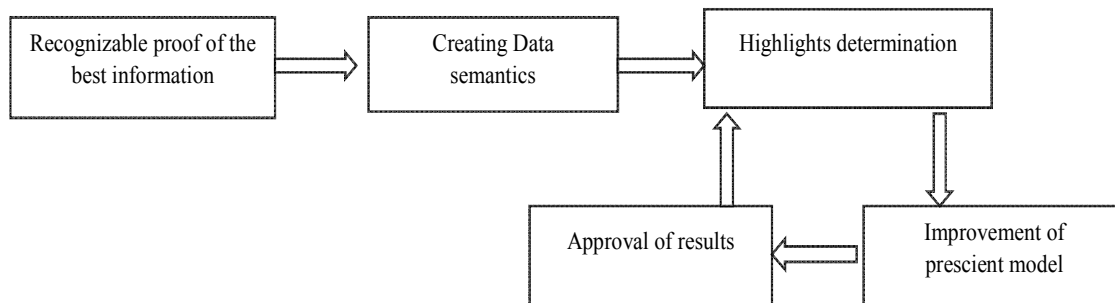


Fig. 1: The phases of churn administration structure

V. PROPOSED WORK

Recognizable proof of customer anticipation to end their authoritative administration is a serious issue. It is identified from a literature overview that many researchers' working in this field has picked to construct their exploration in light of statistic and mining techniques. For a legally binding administration, statistically utilizing the challenging information are the greatest helpful decision for creating a customer agitate forecast philosophy for the accompanying factors: Study of a customer's standard use conduct and checking action, sudden diminished use can be deciphered as a customer's objective to churn. Checking a purchaser's spending conduct; sudden lessened income from a specific one can be deciphered as his/her goal to churn by sectioning them. Fig.2 illustrates the proposed framework in a step by step data flow representation.

Algorithm: 1: A Proposed hybrid algorithm for churn prediction (KSVM+ANN)

Input: Bank marketing data set

Output: Customer churn

Step 1: Load initial dataset

Step 2: Perform the data preprocessing

Step 3: Classification using KSVM

Step 3.1: Split Data set (D_{tot}) into Training data (D_{tr}) and Validation (D_{val})

Step 3.2: Train the data (D_{tr}) using KSVM

Step 3.3: Calculate the distance using Radial function using eq. (2).

Step 3.4: Derive number of classes

Step 3.5: Validate the data (D_{val}) from the derived class

Step 4: Predict the churn from validated data

Step 5: Re-assign the predicted values as Target class for new model

Step 6: Classification using ANN.

Step 6.1: Set hidden layer size

Step 6.2: Set net Type: Multi-Layer Feedforward neural network

Step 6.3: Set hidden neurons: using eq. (1).

Step 6.4: Train the network: $[net, t_r] = (net, inputs, target)$

Step 6.5: Validate the network: $output = net(inputs)$

Step 7: Predict the customer churn from validated data

Step 8: End.

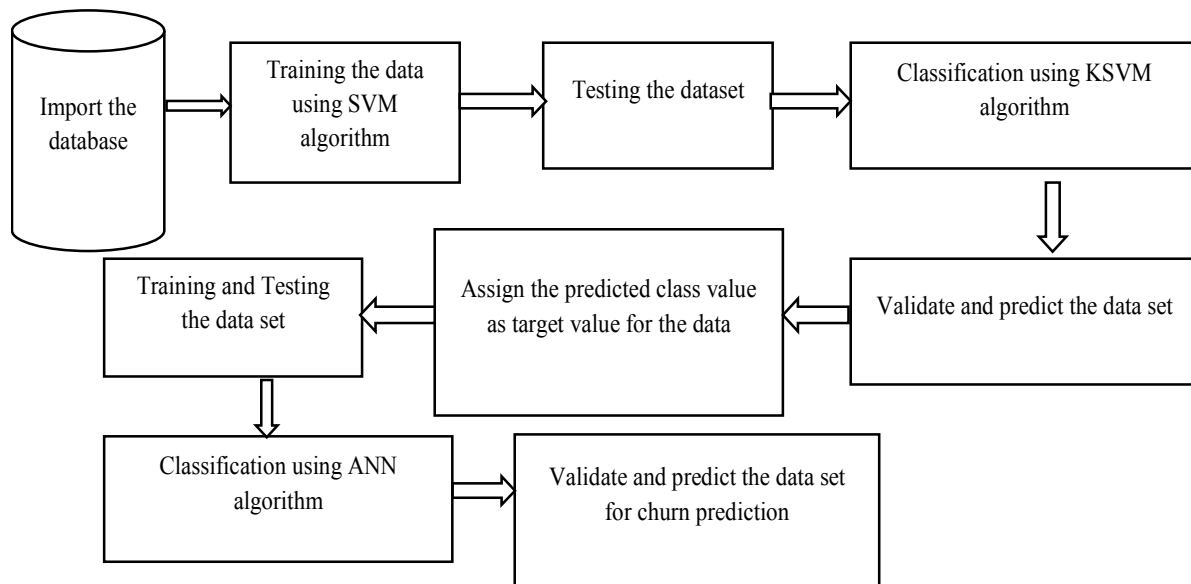


Fig.2: Proposed work

VI. RESULTS AND DISCUSSION

This work is implemented using R Studio version 1.0.136. The input dataset comprises of customer personal details and financial details. It is obtained from the UCI Repository. It consists of 20 attributes. The attributes are namely Name, gender, Age, Marital, Education, profession, Bank, Default, Balance, housing, Month, Duration, Campaign, Pdays, Previous, Pout come, Status, Loan, Contact, Day. Fig.3 tables the metrics used in the comparison with the different algorithms used with the proposed model.

A. Classification Using KSVM Algorithm

KSVM calculation is utilized to group the record among 20 attributes. It underpins both numeric and character sort esteems. "Support Vector Machine" (SVM) is a supervised machine learning calculation which can be employed for mutual arrangement and logistic tests. In any case, it is, for the most part, exploited as a part of characterization issues. With this intention, we have plot each data thing as a fact in n-dimensional space (where n is a number of elements we have) with the estimation of each element being the approximation of a precise facilitate. At that point, we perform an arrangement by finding the hyperplane that separates the two classes exceptionally well (take a gander at the underneath preview). In machine learning, kernel techniques are a class of calculations for design analysis, whose best-known associate is the support vector machine (SVM). Most of the kernel processes are based on convex optimization or eigen problems and are mathematically understandable.

B. Classification Using ANN Algorithm

Feed-Forward Neural Networks and Multinomial Log-Linear Models are used to classify the data. It is an iterative algorithm. A feed-forward neural system is a simulated neural system where associations amongst the units don't surround a coordinated sequence. This is not quite the same as repetitive

neural systems. The feed-forward neural system was the first and ostensibly most straightforward kind of artificial neural system conceived. In this framework, the information moves in only a solitary bearing, forward, from the data center points, through the covered centers (expecting any) and to the yield center points. There are no cycles or circles in the framework. Feed- Forward Neural Networks and Multinomial Log-Linear Models based simulated Neural Network decision the information all the more precisely. But it allows only numeric values.

C. Classification Using KSVM and ANN Algorithm

The adopted model KSVM and ANN showed better accuracy results and even resulted in model improvement when examined the metrics with the traditional classifiers. The hybrid combination was a result of fetching the output produced by the KSVM and running the model with ANN classifier resulting in better model prediction values of churn data in the database acquired.

D. Evaluation Metrics

TABLE I. MODEL EVALUATION

MODELS	SVM	KSVM	ANN	KSVM+ANN	Derivations
Measure	Value				
Sensitivity	0.9432	0.9490	0.9789	0.9897	$TPR = TP / (TP + FN)$
Specificity	0.4167	0.7143	0.5000	0.5000	$SPC = TN / (FP + TN)$
Precision	0.9222	0.9789	0.9789	0.9897	$PPV = TP / (TP + FP)$
Negative Predictive Value	0.5000	0.5000	0.5000	0.5000	$NPV = TN / (TN + FN)$
False Positive Rate	0.5833	0.2857	0.5000	0.5000	$FPR = FP / (FP + TN)$
False Discovery Rate	0.0778	0.0211	0.0211	0.0103	$FDR = FP / (FP + TP)$
False Negative Rate	0.0568	0.0510	0.0211	0.0103	$FNR = FN / (FN + TP)$
Accuracy	0.8800	0.9333	0.9596	0.9798	$ACC = (TP + TN) / (P + N)$
F1 Score	0.9326	0.9637	0.9789	0.9897	$F1 = 2TP / (2TP + FP + FN)$
Matthews Correlation Coefficient	0.3898	0.5636	0.4789	0.4897	$\frac{TP*TN - FP*FN}{\sqrt{((TP+FP)*(TP+FN)*(TN+FP)*(TN+FN))}}$

Fig.3: Results of the metrics

E. Performance of the Model

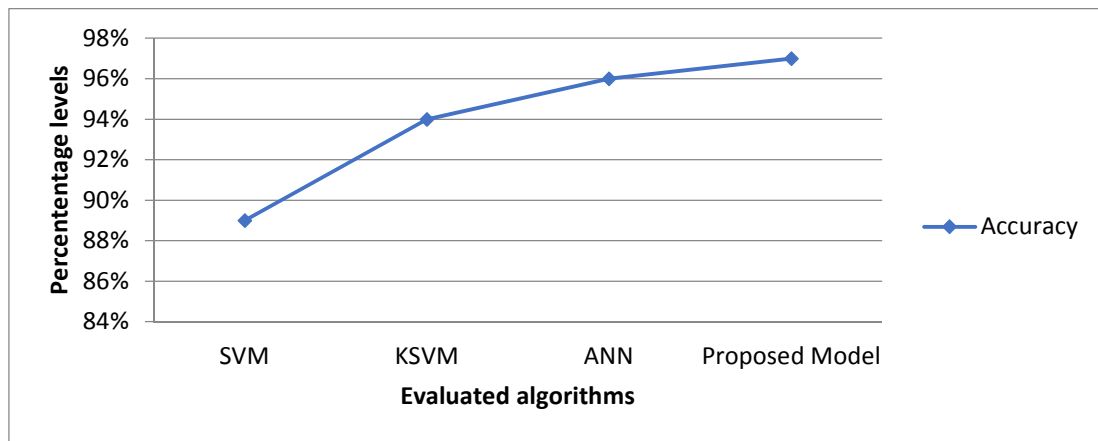


Fig.4: Accuracy of models

VII. CONCLUSION AND FUTURE WORK

Besides adopting ANN and KSVM we proposed a hybrid approach of predicting churn customer in Indian bank sectors. A broad arrangement of regressions affirms that the proposed approach displays 97% of precision and the outcome admitted one of the best in class procedures as far as both runtime and memory utilization [2][10]. Fig.4 depicts the type I accuracy performance of the models. As of late hold, the potential customer has gotten awesome a lot of consideration because of their wide application, has appeared in our outcomes, in spite of the fact that proposed half and half runs quickest, and performs great precision. It expends more memory. The most effective method to enhance the memory utilization of SVM and ANN is an intriguing exploration subject. An intriguing bearing of future work is to stretch out ANN or KSVM to create proficient strategies to anticipate and hold the customers. Furthermore, since the accessible information is developing exponentially, utilizing neural networks and fuzzy techniques to deal with the component decision in huge information is likewise a fascinating work.

REFERENCES

- [1] Zhu, B., Baesens, B., & Vandenbroucke, S. K. (2017). An empirical comparison of techniques for the class imbalance problem in churn prediction. *Information sciences*, Vol.408, pp.84-99.
- [2] Nijhawan, V. K., Madan, M., & Dave, M. (2019). An Analytical Implementation of CART Using RStudio for Churn Prediction. In *Information and Communication Technology for Competitive Strategies*, Springer, Singapore, pp. 109-120.
- [3] Amin, A., Al-Obeidat, F., Shah, B., Adnan, A., Loo, J., & Anwar, S. (2019). Customer churn prediction in telecommunication industry using data certainty. *Journal of Business Research*, Vol. 94, pp.290-301.
- [4] Chandrakala, D. (2016). A Survey on Customer Churn Prediction using Machine Learning Techniques. *International Journal of Computer Applications*, Vol.154, No. 10, pp.13-16.
- [5] De Caigny, A., Coussement, K., & De Bock, K. W. (2018). A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. *European Journal of Operational Research*, Vol. 269, No.2, pp.760-772.
- [6] Mohanty, R., & Sree, C. N. R. (2018). Churn and Non-churn of Customers in Banking Sector Using Extreme Learning Machine. In *Proceedings of the Second International Conference on Computational Intelligence and Informatics*, Springer, Singapore, pp. 51-58.
- [7] Garcia, D. L., Nebot, A., & Vellido, A. (2017). Intelligent data analysis approaches to churn as a business problem: a survey. *Knowledge and Information Systems*, Vol.51, No.3, pp.719-774.
- [8] Ngai, E. W., Xiu, L., & Chau, D. C. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert systems with applications*, Vol. 36, No.2, pp.2592-2602.
- [9] Coussement, K., Lessmann, S., & Verstraeten, G. (2017). A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry. *Decision Support Systems*, Vol.95, pp. 27-36.
- [10] Xiao, J., Jiang, X., He, C., & Teng, G. (2016). Churn prediction in customer relationship management via gmdh-based multiple classifiers ensemble. *IEEE Intelligent Systems*, Vol.31, No.2, pp. 37-44.