

DETERMENT OF CUSTOMER CHURN IN BANKING USING MACHINE LEARNING TECHNIQUES

¹Jannu Neeraja, ²Polasi.Sudhakar

¹M.Tech. Student of Ramachandra College of Engineering, Eluru,A.P,India

²Assistant Professor, Dept of Computer Science & Engineering Ramachandra College of Engineering, Eluru,A.P,India

Abstract: All over the world, in one-of-a-kind sectors churn prediction performs a very vital position in the increase of the organization. For the company's income and earnings patron churn is very harmful. The most vital step to keep away from churn is to become aware of churn and its reason, as a consequence provoke the stopping measures. Now a day's laptop mastering performs a fundamental function to get rid of this problem. The goal of this paper is to predict the churn in banking sectors, with the aid of the use of nicely regarded laptop mastering methods like aid vector laptop (SVM) and Naive-Bayes algorithm. The classification mannequin is constructed via inspecting historic records and then making use of the prediction mannequin primarily based on the analysis. The described mannequin has a prediction fee of 94.63% for an experimental dataset containing purchaser records of a worldwide bank, gathered from Kaggle

Keywords: Naïve-Bayes algorithm, Machine Learning (ML), Support Vector Machine (SVM)

I. INTRODUCTION

To get and hold the loyal client for each enterprise is a huge challenge. Correct prediction about a client is going to churn or no longer and then efficaciously convincing him to remain with that enterprise can enlarge the income of that company. Therefore, predicting patron churn, i.e. if a purchaser is about to go away for a higher service, that is an vital phase for examining the client behaviour. The churn mannequin is a illustration of quite a number calculations that are constructed on current historic data. The patron churn can be described in different approaches also, like low switching cost, deregulation motivates a client to exchange the sector. The churn is additionally categorized into two: voluntary and involuntary churn . Voluntary churn is described as the termination of offerings by means of the consumer itself, whereas involuntary churn is described as the termination of offerings with the aid of the financial institution for fraud, non-payment services.

The client churn is very volatile if it is no longer managed carefully, it might also deliver a organisation to its knees for negative preservation services. Cost of patron churn additionally consists of loss of revenue, profit. Previous case find out about has proven that the value of retaining a new consumer is greater than the value of keeping the historical one .There are a number banks who are struggling with this consumer churn problem. So the most described way to deal with this hassle is growing a predictive mannequin that can reliably and without difficulty become aware of the viable churning. In the latest previous the most regularly used method is facts mining to boost such fashions with first-class consequences . Day with the aid of day when this churn prediction trouble receives importance, lots extra lookup efforts are generated toward enhancing churn prediction rates.

Most often used facets that have been used earlier consists of deposit score, geography, has deposit card or not, is lively member, estimated profits. Due to enterprise privateness and coverage it is tough and use public dataset for churn prediction. Tsai and Lu before used two exceptional hybrid fashions that can predict consumer churn. To enlarge the prediction charge Kechadi and Buckley used attribute derivation method.

In this paper, a new subset of aspects has been introduced in order to enhance the prediction accuracy. Naïve Bayes classifier and SVM have been used for prediction and the outcomes are compared.This paper has been geared up into 5 sections. Section-2 presents small print of dataset and comparison measures, Section-3 demonstrates the proposed device and its methodology. Section-4 gives evaluation of results. Finally, Section-5 concludes this study.

II. METHODOLOGY

1. Dataset

In this Study, a financial institution fact is regarded the place a big quantity of clients are leaving the bank. Almost ten thousand documents of the bank, accrued from Kaggle repository, are going to assist the mannequin to look at and predict which of the clients are about to depart the financial institution soon. To take a look at and consider the points the whole dataset sliced into two subsets, education and checking out dataset. Training dataset can be used to outline the statistical mannequin and the trying out dataset can be used to predict the end result and calculation of accuracy metrics for finding out the mannequin accuracy. For validation of the mannequin accuracy of the classifier is calculated on the foundation Confusion Matrix.

True positive (TP): correctly predicted positive cases.

False negative (FN): positive cases wrongly predicted as negative.

False positive (FP): negative cases wrongly predicted as positive.

True negative (TN): negative cases correctly predicted.

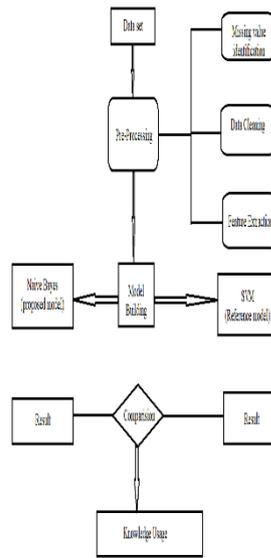


Fig.1. Flow Chart

1. DATA SAMPLING

The dataset set containing ten thousand clients' records in 14 columns, these are customer id, surname, credit score score, gender, age, balance, geography, has credit score card, Tenure, Num of Product, Is Active Member, Estimated Salary and exited (the cost suggests 1 when the purchaser churned).

1. PRE-PROCESSING

In this section one-of-a-kind preprocessing strategies like managing the lacking values, statistics cleansing and characteristic extraction system have been performed. To become aware of the lacking values in the dataset imputation method will be used to impute the clean and null values. Noisy data, inappropriate attributes are removed. Those attributes that are now not so tons necessary are removed for mannequin constructing. Finally for figuring out the overall performance of predictive fashions, characteristic extraction performs an essential position for right prediction. Here are some vital elements with description that can be beneficial for mannequin building as proven in Fig 1.

- Age: the age of the customer.
- Balance: the whole financial institution stability of the customer.
- Is Active Member: how lively the consumer is/was in his carrier period.
- Tenure: described via the size of carrier or duration, from the beginning date to termination date. Plays an essential position for prediction of churn.

In Fig 2, container plots of some essential attributes are given. When it comes to the distribution of all statistics factors over mean, field plots are used to perceive the median and additionally respective traits in a well-structured manner

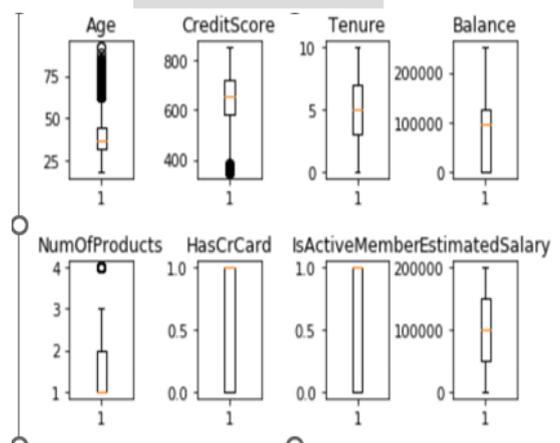


Fig 2. Box-plots of Attributes

2. CLASSIFICATION

In this study, the probabilistic algorithm Naïve Bayes is used on the chosen function set to improve the model. Then the end result is in contrast with the most extensively used algorithm i.e. SVM. The two methods are temporarily mentioned in the subsequent section.

1. Support Vector Machine (SVM)

SVM is a frequent nonlinear method and a theoretically effective device for the classification of all kinds of datasets. The Support Vector Machine (SVM) is an algorithm for supervised computing device studying that can be used for troubles with classification or regression. However, it is basically used for classification problems. After presenting an SVM mannequin series of named coaching information for every category, they are capable to categorize the new class. Enter Support Vector Machines (SVM): a easy and nice classification algorithm that performs very nicely with a constrained quantity of data. In every type the closest factors are denoted as guide vectors. If different factors are left out in the coaching set the measured judgment boundary stays the same. We extract the quintessential features and section them into information for education and testing. Eighty percentage of the facts is used for coaching and the closing 20 percentage for testing. The SVM mannequin was once developed the use of the NumPy library. An α (0.0001) is the studying charge and is set to $1 / \text{epochs}$ via the regularization parameter λ . Hence the regularizing cost reduces the quantity of will increase in epochs. After classifying the dataset, we get the poor records from which to observe the proposed model. Various ML strategies like linear regression, Decision tree, Random Forest, and Support Vector Machine (SVM) are used to assemble the classifier. Where is a sturdy entire approach for constructing the mannequin amongst all of SVM? SVM is used as a classifier in this methodology. However, it is basically used for classification problems. In the SVM algorithm, we map every information aspect as a factor in n-dimensional area (where n is the quantity of aspects you have) with every function being the price of a precise coordinate. Then, we elevate out classification via discovering the hyperplane that separates the two training very well. In Fig 3, SVM is given.

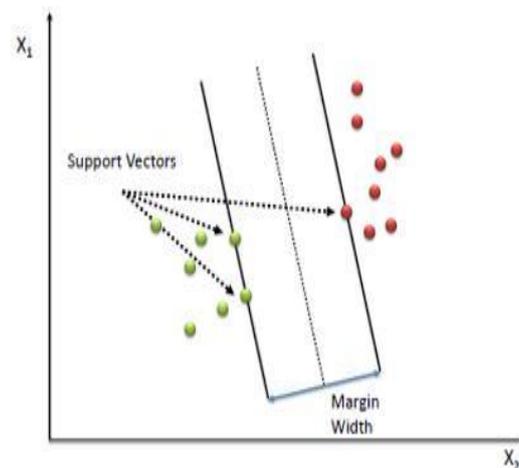


Fig.3.Supportvector Machine(SVM)

2. Naïve Bayes Algorithm:

This is based totally on the Bayesian theorem. The strategy immediately analyses the relationship between every characteristic as nicely as the category of every occasion primarily based on conditional probability. Assume that X is denoted as the vector of occasion and additionally a random variable C denoted This as the type of an instance. So now determine that X is a precise occasion and C is a precise class. Now the education procedure suggests that the computation of the chance of each classification is described through how many instances the category takes place in the education dataset. This is recognized as the prior chance $P(C=c)$. As properly as the chance of the instance x given c is additionally computed by way of this algorithm. The benefit of this algorithm is that Naïve Bayes has carried out suitable end result when the assumption and the viewed classification is violated. So, the Bayes components is as follows:

$$p(X = x) = \frac{P(C=c)\pi_i P(x_i|C=c)}{P(X=x)} \quad (1) \text{ The likelihood of an occasion (x) and a type c can be computed via the mixture of prior chance and attribute density function.}$$

3.ASSESSMENT OF CLASSIFICATION RESULTS

The mannequin is evaluated the usage of two special desktop studying algorithms SVM and Naïve-Bayes algorithm by way of the usage of a take a look at information measurement 2000, which is 20% of the complete dataset. The overall performance of the classifiers evaluated in phrases of sensitivity, specificity and universal accuracy as proven in [Table 1]

Based on the all pairs of True Positive (TP) and False Positive (FP) the receiver working attribute curve (ROC) curves [12, 13] have been plotted as proven in Fig.4. the place X axis represents the false churn price and Y axis represents the real churn rate. Each ROC curve represents the genuine pair of prediction charge (FP, TP). The ROC curve which is nearer to the left pinnacle nook represents higher prediction rates.

The distinction between the Naïve Bayes and SVM has proven from the top curves. It is additionally proven that the sampling charge CHURN/ACTIVE decreases the quantity of ACTIVE will increase via the Naïve Bayes model. This learn about confirmed that Naïve Bayes probabilistic classifier reap greater boom than the SVM.

The reap of statistics is measured via the statistics concept that additionally measures the data content material on the attribute. The attribute we use to construct the mannequin is all about excessive facts gain. Depending on the statistics acquire of the attribute we use the pinnacle ranked attribute for mannequin constructing purpose. It is described that the chosen attribute has the easiest prediction charge via the Naïve Bayes model.

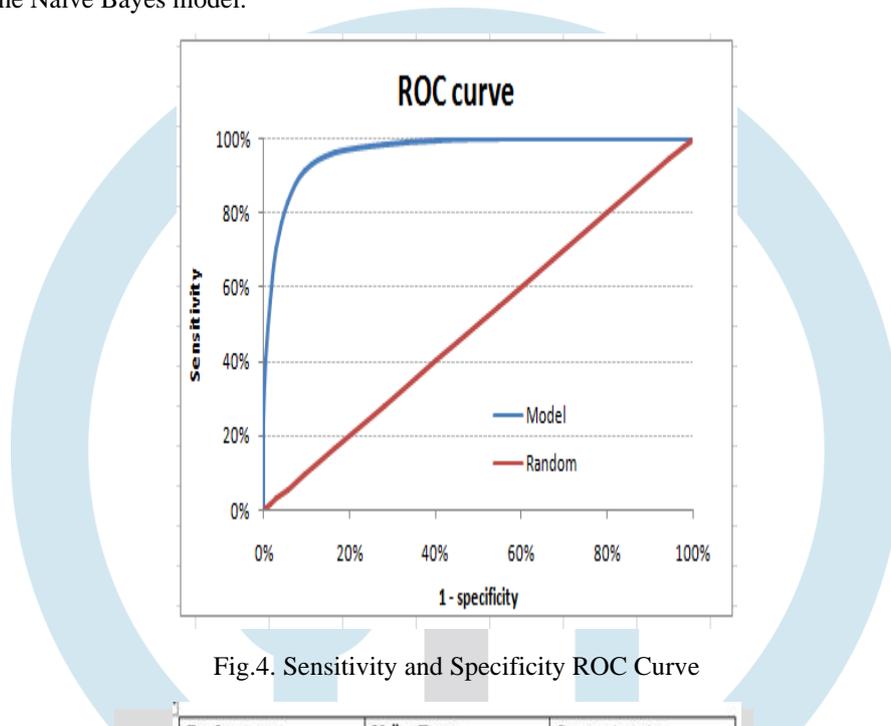


Fig.4. Sensitivity and Specificity ROC Curve

Performance Measure	Naïve Bayes	Support vector machine (SVM)
Accuracy	94.63%	92.25%
Sensitivity	95.54%	94%
Specificity	94.25%	92%

Table 1: Performance Measurement

III. CONCLUSION

This paper addressed a computerized churn prediction approach that is capable to predict the churn in banking sectors, by using the use of properly regarded computer studying strategies like aid vector computer (SVM) and Naive-Bayes algorithm. With Naive-Bayes the algorithm the mannequin is in a position to reap an accuracy of 94.63%, when examined with an global financial institution dataset accessible at Kaggel repository. Thus, it can be used for the actual time application.

IV. REFERENCES

- [1] Seo D, Ranganathan C, Babad Y. Two-level mannequin of client retention in the US cell telecommunications provider market. *Telecommun Policy* 2008; 32 (3):182–96.
- [2] G. M. Weiss, "Mining with rarity: a unifying framework," *ACM SIGKDD Explorations Newsletter*, vol. 6, no. 1, pp. 7–19, 2004.
- [3] J. Burez and D. Van den Poel, "Handling type imbalance in consumer churn prediction," *Expert Systems with Applications*, vol. 36, no. 3, pp. 4626–4636, 2009.
- [4] M. T. K. a. B. B. Huang, "Customer churn prediction in telecommunications," *Expert Systems with Applications*, vol. 39, no. 1, pp. 1414-1425, 2012.

- [5] W.-H. a. C. K. C. a. Y. X. Au, "A novel evolutionary statistics mining algorithm with functions to churn prediction," IEEE transactions on evolutionary computation, vol. 7, no. 6, pp.532-545,2003.
- [6]<https://www.kaggle.com/filippoo/deep-learning-az-ann>
- [7] Chu, B. H., Tsai, M. S., and Ho, C. S., "Towards a hybrid information mining mannequin for client retention", Knowledge-Based Systems, 20, 2007, pp. 703–718.
- [8] Shui Hua Han a, ShuiXiuLu a, Stephen C.H. Leung., "Segmentation of telecom clients primarily based on client price via choice tree model", Expert Systems with Applications, 39, 2012,3964–3973
- [9] VivekBhambri, "Data Mining as a Tool to Predict Churn Behaviour of Customers", GE-International Journal of Management Research (IJMR).
- [10] S. Gupta, D. Hanssens, B. Hardie, W. Kahn, V. Kumar, N. Lin, N. Ravishanker, and N. R. S. Sriram, "Modeling Customer Lifetime Value," J. Serv. Res., vol. 9, no. 2, pp. 139–155, Nov.2006.
- [11] K. Muata, and O. Bryson, Evaluation of Decision Trees: AMulti Criteria Approach, Computers and Operational Research, 31, 1933- 1945, 200
- [12] P. Mohapatra, S. Chakravarty, P. K. Dash, Microarray scientific facts classification the usage of kernel ridge regression and modified cat swarm optimization based totally gene decision system, Swarm and Evolutionary Computation, 28, 144-160, 2016.

