

EXAMINATION AND DETERMINATION OF CUSTOMERS' CHURNS USING DATA MINING MODELS IN THE BANKING INDUSTRY

¹R.Sagar, ²Y.Susheela, ³A.Sowjanya, ⁴Katukojula Shivaprasad

^{1,3}Assistant Professor, ²Associate Professor, ⁴UG Student, ^{1,2,3,4}Department of CSE, Vaageswari College of Engineering, Karimnagar, Telangana, India

ABSTRACT

A fresh approach to studying and forecasting client churn has been put forth. The technique uses a data mining paradigm in the banking sector. This was brought about by the fact that there are an estimated 1.5 million churning consumers per year, a number that is continuously growing. Churn customer prediction is the process of determining whether or not a customer will leave a business. One technique to predict customer attrition is to use a classification strategy from data mining that creates a machine learning model. This study tested five different categorization algorithms using a dataset of 57 variables. Multiple comparison experiments between various classes were carried out. With a 50:50 comparison, the Support Vector Machine (SVM) was used. At an Indonesian private bank, class sampling data is the most effective tool for predicting client attrition. The results of this modeling can be used by businesses to take strategic measures to avoid client attrition.

Keywords—customer churn, prediction, data mining, classification, machine learning

INTRODUCTION

One of the largest banks in Indonesia, XYZ Bank, has thousands of thousands of customers who must be treated well in order for them to continue using the services offered by the business. Businesses have understood that they must work to retain current customers in addition to acquiring new ones since, if current customers churn, the number of clients would decrease if there are no additional new consumers. Around 1.5 million customers at our case study (XYZ Bank) leave each year, and that number is rising. Despite the fact that it may contribute to the deterioration of new customers, to get new clients fees 5 to six instances higher than preserving current customers. Some techniques can be executed to protect ancient customers, which is to predict clients who will churn. Predicting churnclients goals to pick out potential churnclients primarily based on previous dataand preceding behavior so that incentives can be provided to survive. Data evaluation can be described as an in-depth examination of the means and necessary values handy in the records to discover necessary data the use of precise strategies and techniques [3]. One approachthat can be used is information mining techniques. Some preceding research [1],[4],[5],[6] many have shown that statistics mining methods can be used to predict churn customers. The cause of this find out about is to attain the first-class facts mining studying mannequin thatcan be carried out by way of XYZ Bank to stop clients from leaving them.

LITEARTURE SURVEY

Customer churn analysis in banking sector using data miningtechniques

AUTHORS: Oyeniyi, A., & Adeyemo

Customer churn has become a major problem within a customer centred banking industry and banks have always tried to track customer interaction with the company, in order to detect early warning signs in customer's behaviour such as reduced transactions, account status dormancy and take steps to prevent churn. This paper presents a data mining model that can be used to predict which customers are most likely to churn (or switch banks). The study used real-life customer records provided by a major Nigerian bank. The raw data was cleaned, pre-processed and then analysed using WEKA, a data mining software tool for knowledge analysis. Simple K-Means was used for the clustering phase while a rule-based algorithm, JRip was used for the rule generation phase. The results obtained showed that the methods used can determine patterns in customer behaviours and help banks to identify likely churners and hence develop customer retention modalities. The regulatory framework within which financial institutions and insurance firms

operate require their interaction with customers to be tracked, recorded, stored in Customer Relationship Management (CRM) databases, and then data mine the information in a way that increases customer relations, average revenue per unit (ARPU) and decrease churn rate. According to, churn has an equal or greater impact on Customer Lifetime Value (CLTV) when compared to one of the most regarded Key Performance Indicator (KPI's) such as Average Revenue Per User (ARPU). As one of the biggest destructors of enterprise value, it has become one of the top issues for the banking industry. Customer churn prediction is aimed at determining customers who are at risk of leaving, and whether such customers are worth retaining.

Analytical model of customer churn based on Bayesian network

AUTHORS: Peng, S., Xin, G., Yunpeng, Z., & Ziyang, W

A customer churn analytical model based Bayesian network is built for prediction of customer churn. We propose Bayesian Network approaches to predict churn motivation, mining the result in churn characters in order to help decision making manager formulate corresponding detainment strategy. Experimental results show that classification performance of both methods is resultful. Customer churn is a big problem in marketing for long time. Companies have become aware that they should put much effort not only trying to convince customers to sign contracts, but also to retain existing clients. On one hand, customer churn may decrease the sale. On the other hand, it may lead to the reduction of new customers. Moreover, the cost of gaining a new customer is five to six times of retaining an old customer. In terms of companies, it's very important to build a well-defined model for customer churn which can explain who and why is likely to churn. Resultful approaches must be found in order to help decision-making manager formulate corresponding detainment strategy. There are two kinds of methods presented by scholars to analyze customer churn. One is traditional classification method, such as Decision Tree, Logistic Regression, Naive BN and Clustering Analysis. The other is artificial intelligence method, such as Artificial Neural Network (ANN), Self-organizing Map (SOM) and Evolutionary Learning. This paper adopts Bayesian networks to build a model of customer churn analysis which can analyze the probability of churn factors and dissect the customer behavior to provide decision basis for marketing.

Designing of customer and employee churn prediction model based on data mining method and neural predictor

AUTHORS: Dolatabadi, S. H., & Keynia, F

In recent years due to increased competition between companies in the services sector, predict churn customer in order to retain customers is so important. The impact of brand loyalty and customer churn in an organization as well as the difficulty of attracting a new customer per lost customer is very painful for organizations. Obtaining a predictive model customer behaviour to plan for and deal with such cases, can be very helpful. Employee churn or loss of staff will be close to the customer churn, but the impact of losing a major customer for organization certainly will be more painful (because organization do not have physical sense to losing their employees) while the consequences of finding well employees instead of missed employees, As well as the cost of in-service training that should be given to new employees could be one of the issues that each organization would be sensitive to losing its human resources.

A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector

AUTHORS : Irfan Ullah, Basit Raza, Ahmad Kamran Malik , Muhammad Imran, Saif Ul Islam , And Sung Won Kim

In the telecom sector, a huge volume of data is being generated on a daily basis due to a vast client base. Decision makers and business analysts emphasized that attaining new customers is costlier than retaining the existing ones. Business analysts and customer relationship management (CRM) analyzers need to know the reasons for churn customers, as well as, behavior patterns from the existing churn customers' data. This paper proposes a churn prediction model that uses classification, as well as, clustering techniques to identify the churn customers and provides the factors behind the churning of customers in the telecom sector. Feature selection is performed by using information gain and correlation attribute ranking filter. The proposed model first classifies churn customers data using classification algorithms, in which the Random

Forest (RF) algorithm performed well with 88.63% correctly classified instances. Creating effective retention policies is an essential task of the CRM to prevent churners. After classification, the proposed model segments the churning customer's data by categorizing the churn customers in groups using cosine similarity to provide group-based retention offers. This paper also identified churn factors that are essential in determining the root causes of churn. By knowing the significant churn factors from customers' data, CRM can improve productivity, recommend relevant promotions to the group of likely churn customers based on similar behavior patterns, and excessively improve marketing campaigns of the company. The proposed churn prediction model is evaluated using metrics, such as accuracy, precision, recall, f-measure, and receiving operating characteristics (ROC) area. The results reveal that our proposed churn prediction model produced better churn classification using the RF algorithm and customer profiling using k-means clustering. Furthermore, it also provides factors behind the churning of churn customers through the rules generated by using the attribute-selected classifier algorithm.

PROPOSED WORK

In this Proposed device makes use of deductive technique and the kind of lookup is case learn about lookup and experimental research. The test used to be carried out by means of developing a facts mining getting to know mannequin that pursuits to predict clients who will churn. From these troubles the lookup query located is "what is the quality classification mannequin that can be used to predict churn customers, thereby lowering the threat of clients going to Bank XYZ?". All gaining knowledge of fashions produced are then evaluated to get the satisfactory gaining knowledge of mannequin that great suits the case to be completed. For the lookup phase, this find out about makes use of CRISP-DM as a framework. Data instruction refers to the formation of dataset training, trying out datasets and validation statistics referring to the instruction of stability data, transactions and demographic statistics that will be used as enter facts from the mannequin to be made.

IMPLEMENTATION

The User can register the first. While registering he required a valid user email and mobile for further communications. Once the user register then admin can activate the customer. Once admin activated the users then user can login into our system. The user can see the dataset of banking customers. In the dataset there is 10000 records are there to find which customer can leave the bank. And we took these are the attributes of the users CustomerId, Surname, CreditScore, Geography, Gender, Age, Tenure, Balance, NumOfProducts, HasCrCard, IsActiveMember, EstimatedSalary, Exited. By using this attributes user can perform the model execution by specified attributes.

.Admin:

Admin can login with his credentials. Once he login he can activate the users. The activated user only login in our applications. The admin can set the training and testing data for the project dynamically to the code. Here 1/3 of data split into training and testing. Admin can view all the models training and testing accuracy and training accuracy. After that the admin can see the confusion matrix of the algorithms.

Data Preprocess:

Pre-processing refers to the transformations applied to our data before feeding it to the algorithm. Data Preprocessing is a technique that is used to convert the raw data into a clean data set. In other words, whenever the data is gathered from different sources it is collected in raw format which is not feasible for the analysis.

For achieving better results from the applied model in Machine Learning projects the format of the data has to be in a proper manner. Some specified Machine Learning model needs information in a specified format, for example, Random Forest algorithm does not support null values, therefore to execute random forest algorithm null values have to be managed from the original raw data set. Another aspect is that data set should be formatted in such a way that more than one Machine Learning and Deep Learning algorithms are executed in one data set, and best out of them is chosen.

Models:

At this stage a classification model will be established in accordance with the criteria and settings previously explained. There are five classification methods that will be used, namely Decision Tree, Neural Network, Support Vector Machine (SVM), Naïve Bayes and Logistic Regression. Recall is something important to compare because the recall value is the percentage of success of the model in predicting the true customer churn which is actually the

RESULTS AND DISCUSSION

churn of all the customers who actually churn. In other words, this value is the value that represents the success of the company in making the customers who initially want to go to not go. Precision is the correct percentage of the model predicting the true customer churn which is actually compared to the total of all customer customers predicted by churn. This will have an effect in calculating losses incurred by the company if it follows up on customers who are wrongly predicted

RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCreditCard	IsActiveMember	EstimatedSalary	Exited
0	1	19654002	819	France	Female	42	2	0.00	1	1	1	101342.00	1
1	2	19647311	638	Spain	Female	41	1	83807.80	1	0	1	112542.00	0
2	2	19619854	502	France	Female	42	0	158863.00	2	1	0	112003.57	1
3	4	19301254	699	France	Female	39	1	0.00	2	0	0	93026.00	0
4	5	19379189	850	Spain	Female	40	2	123110.02	1	1	1	78984.10	0
5	9	19374012	645	Spain	Male	44	8	113795.70	2	1	0	140766.71	1
6	7	19302037	822	France	Male	30	7	0.00	2	1	1	10802.00	0
7	8	19686148	576	Germany	Female	29	4	173046.74	4	1	0	119348.00	1
8	9	19702385	501	France	Male	44	4	142201.07	2	0	1	104940.50	0

Fig 4.1 View Dataset

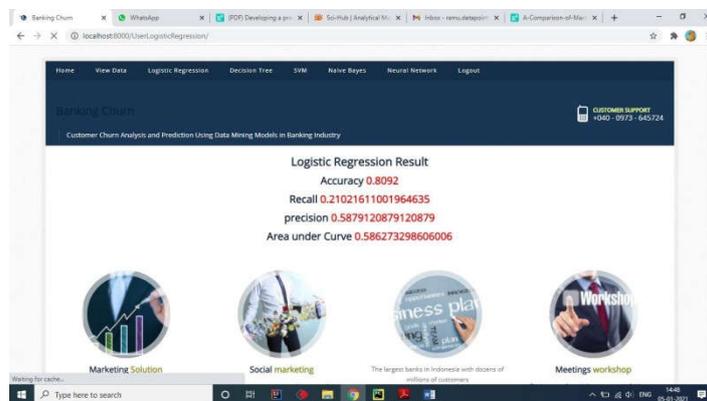


Fig 4.2 Logistic Regression Results:

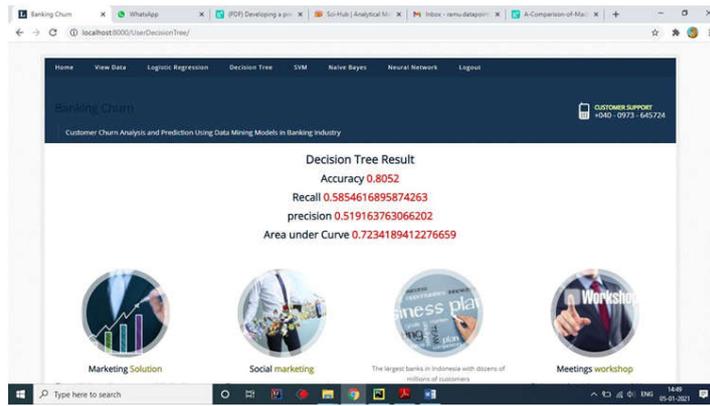


Fig 4.3 Decision Tree Result:

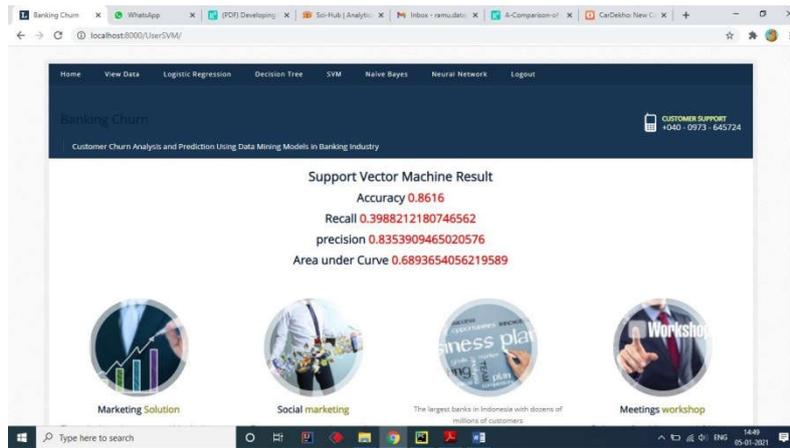


Fig 4.4 SVM Result:

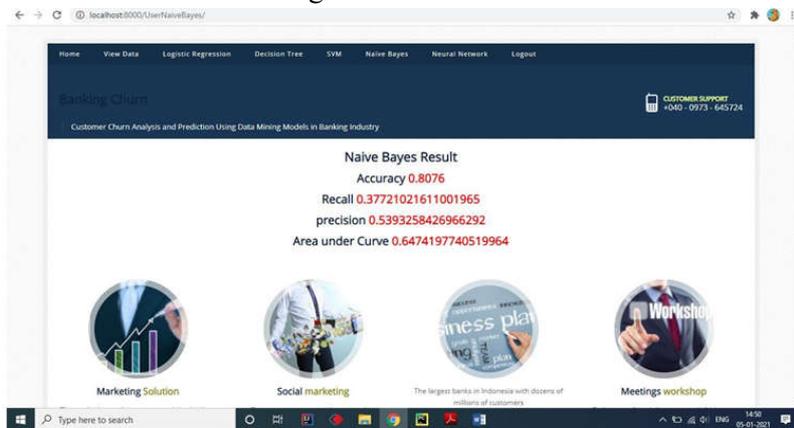


Fig 4.5 Naïve bayes Result:

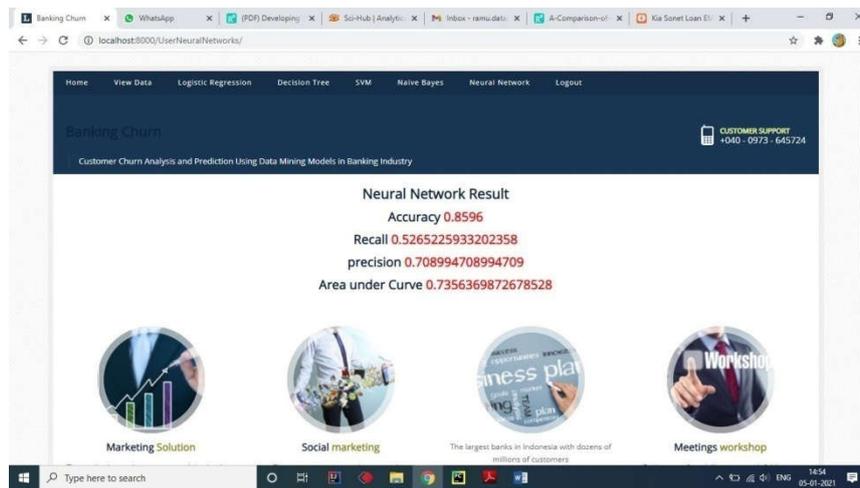


Fig 4.6 Neural Network Results

CONCLUSION

Transaction, amount of (Electronic Data Capture) EDC (Electronic Data Capture)

It is confirmed that information mining can be utilized to forecast customer attrition in the banking industry. The results of modeling are significantly influenced by the vast range of information samples utilized for learning. The memory results are significantly influenced by the variety of interclass comparisons, with the analysis of the 50:50 statistics producing a greater recall rate (on average 70%) than the other two options. In order to gather information, about 15.949 samples of data were used, with around 7.975 examples of each category. If the distribution of the facts is excessively imbalanced, accuracy levels cannot be fully used as a reference for evaluation. The fantastic mannequin is the mannequin with the best earnings value, particularly the 50:50 SVM sampling mannequin with a earnings fee of 456 billion with loss and gain calculations such as Table 5.6 with the 5 most huge attributes is vintage, extent of EDC transaction, common stability in one month and age. This is in line with the lookup of Dolatabadi et al. (2017) which received SVM as modeling with the fine accuracy in its research, however Logistic Regression is additionally well worth thinking about due to the fact it effects in smaller losses.

REFERENCES

1. Oyeniyi, A., & Adeyemo, A. (2015). Customer churn analysis in banking sector using data mining techniques. *AfricanJournal of Computing & ICT Vol 8*, 165- 174.
2. Peng, S., Xin, G., Yunpeng, Z., & Ziyang, W. (2013). Analytical model of customer churn based on bayesian network. *Ninth International Conference on Computational Intelligence and Security* (pp. 269-271). IEEE.
3. Banarescu, A. (2015). Detecting and preventing fraud with data nalytics. *ScienceDirect*, 2.
4. Dolatabadi, S. H., & Keynia, F. (2017). Designing of customer and employee churn prediction model based on data mining method and neural predictor. *The 2nd International Conference on Computer and Communication Systems* (pp. 74-77). IEEE.
5. Keramati, A., Ghaneei, H., & Mohammad Mirmohammadi, S. (2016). Developing a prediction model for customer churn from electronic banking services using data mining. *Financial Innovation*, 2-10.
6. Zoric, A. B. (2016). Predicting customer churn in banking industry using neural networks. *Interdisciplinary Description of Complex Systems*, 116- 124.
7. Chitra, K., & Subashini, B. (2011). Customer retention in banking sector using predictive data mining technique. *International Conference on Information Technology. ICIT*.