Customer Churn Analysis and Prediction Using Data Mining Models in Banking Industry

[1]Prof. Umavane Kanchan , [2]Mr. Vishal V.Shigwan, [3]Mr. Pratik P.Gondhali,
[4]Miss. Gauri S. Kamble.
[1]Asst.Prof.,[2,3,4]UG Student,[1,2,3,4]Computer Engg. Dept. Shivajirao S. Jondhle College of Engineering & Technology, Asangaon, Maharashtra, India.
[1]kanchanumavane2020@gmail.com, [2]vishushigwan02101998@gmail.com,
[3]pratikgondhali50@gmail.com, [4]gaurikamble2000@gmail.com

Abstract - A novel approach to analyzing and forecasting client attrition has been put forth. The approach makes use of a data mining paradigm in the financial sector. This was motivated by the approximately 1.5 million consumers who churn each year, a number that is rising. Churn customer prediction is the process of determining whether a customer will stay with the business or not. Using a categorization strategy from data mining that generates a machine learning model is one way to predict this customer attrition. Using a datasets of 57 variables, this study investigated 5 alternative classification techniques. Several comparisons be the most effective technique for predicting customer attrition at an Indonesian private bank is Support Vector Machine (SVM) with a comparison of 50:50 Class sampling data. The organization can use the outcomes of this modelling to inform its strategic client retention initiatives Between various classes were used in experiments[1].

Keywords- customer churn, data mining , Support Vector Machine (SVM), machine learing.

I. INTRODUCTION

At this case study, there are around 1.5 million churn customer in a year and increasing every year Although it can have an impact on the decline of new customers, to get new customers costs five to six times greater than retaining existing customers.

Some technique's can be use to defend old customers, which is to predict customers who will churn. Some previous research may have shown that data mining techniques can be used to predict churn customers. The purpose of this study is to obtain the best data mining learning model that can be implemented by Certain Bank to prevent customers from leaving them.[1].

Customer Churn - In the definition of banking, a churn customer can be defined as the person who closes all of his accounts and stops doing business with the bank. Churn customers not only can result in depreciation of funds but also can reduce company profits and other negative impacts on the company's operations[8].

II. AIMS AND OBJECTIVE

a) Aim

The purpose of the project is to create model which can successfully predict churn customer based on varied data mining models. Which is implemented to reduce negative reviews on the bank and depreciation's of funds.

b) Objective

1. The process of transforming a user-centered description of the input into a computer-based system in input design. Its design is crucial to preventing mistakes in the data entry process and providing management with clear instructions for receiving the right information from the computerized system.

2. It is accomplished by designing interfaces that are easy for users to use when entering big amounts of data. The purpose of input design is to make data entering simpler and error-free. The data entering panel is created such that any data manipulations are possible. Moreover, it offers record viewing capabilities.

III. LITERATURE SURVEY

PAPER 1: Customer Churn Prediction Analysis

AUTHORS: Fenil Shah and Mrugendrasinh L. Rahevar

This paper demonstrates how the prediction of customer turnover has become a significant study topic due to the growth of the industry. The goal is to describe the steps involved in creating a churn prediction model, its usage and

 DOI : 10.5281/zenodo.7856491

causes, obstacles and problems encountered during model development, and methods for reducing the churn rate. Suppliers are required to put forth extra effort to meet expectations and prevent a commotion[4].

PAPER 2: 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

This paper proposes a churn prediction model that uses classification and clustering techniques to identify churn customers and provide the factors behind the churning of customers in the telecom sector. The proposed model is evaluated using metrics such as accuracy, precision, recall, f-measure, and receiving operating characteristics (ROC) area. The results show that the proposed model produced better churn classification using the RF algorithm and customer profiling using k-means clustering. It also provides factors behind the churning of churn customers through the rules generated by the attribute-selected classifier algorithm[2].

PAPER 3: Customer Churn Prediction Modelling Based on Behavioural Patterns Analysis using Deep Learning

AUTHORS: S. Agrawal, A. Das, A. Gaikwad and S. Dhage

The customer user records data given by the firm will be used to inform the technique used in this article. It will be necessary to tidy up this raw data before using it. To find the variation in the data and use the appropriate procedures to normalize it, a thorough study is needed. Following this, parameters that will be chosen to make up the feature set used to train a model will be employed. A multi-layered Artificial Neural Network created specifically for this issue is then given this processed data. The model is validated using validation data after it has been trained using training data, and it is then tested using the testing set. As a result, the model in issue would be quantified using a percent accuracy metric. A correlation graph will be produced to identify the key elements that contributed most to the churn. Numeric parameters will also be the subject of micro analysis in an effort to identify patterns and assess how well they fit into general churn trends[3].

IV. EXISTING SYSTEM

The Previous research said that the use of data mining can help predict customer churn. These studies use various methods in learning. Oyeniyi & Adeyemo in 2015 examined the use of customer demographic data and customer transaction data to detect churners using the k- Means method and JRIP algorithm. This study only uses 500 data with only four attributes. This is because this study only focuses on customers who are still carrying out transaction activities within a span of two months before the customer closes their account. This study managed to group customers into five groups using the k-Means algorithm and then processed using the JRIP algorithm which produced an analysis model and evaluated using 10-fold confusion matrix and cross validation[6].

V. COMPARATIVE STUDY

Table No1: Comparative study

Sr. No.	Author	Project Title	Publication	Technology	Purpose
1.	Fenil Shah and Mrugendrasinh L. Rahevar	Customer churn analysis in banking sector using data mining techniques	ResearchGate, 2018	CRM	To find various prediction model to analyise customer charn.
2.	Irfan Ullah, Basit Raza, Ahmad Kamran Malik , Muhammad Imran,Saif Ul Islam , And Sung Won Kim	A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector	IEEE, 2019	Random Forest	It build for prediction of customer churn.
3.	S. Agrawal, A. Das, A. Gaikwad and S. Dhage	Customer Churn Prediction Modelling Based on Behavioural Patterns Analysis using Deep Learning	IEEE, 2018	Neural Network	It find the churn factors and churn analysis

 DOI : 10.5281/zenodo.7856491

VI. PROBLEM STATEMENT

The classification method for predicting churn customers is also influenced by customer product ownership data. This data is used because the tendency0 of customers to leave can be seen from the number of products they have or the number of products that suddenly change as evidenced by previous research.

VII. PROPOSED SYSTEM

The deductive technique and case study and experimental research are the types of research used in the proposed system. A data mining learning model that seeks to anticipate clients who will churn was developed for the trial. What is the best classification model that can be used to anticipate customer turnover, hence lowering the risk of customers going to Bank XYZ? is the research question derived from these issues. The best learning model that best fits the case to be finished is chosen after all the produced learning models have been examined. This study employs CRISP-DM as a framework for the research phase. Data preparation is the process of creating training, testing, and validation datasets. It also refers to the preparation of balance, transaction, and demographic data that will be utilized as input for the model that will be created[1].

VIII. ALGORITHM

The Algorithm for Translating Plain Text to Braille:

Step.1: Read from a CSV file using pandas library

```
path = settings.MEDIA_ROOT + "\\" +
"Churn_Modelling.csv"
   data = pd.read_csv(path, delimiter=',')
   data = data.drop(['CustomerId', 'Surname',
'RowNumber'], axis=1)
   x = data.iloc[:, 0:10]
   y = data.iloc[:, 10]
   x = pd.get_dummies
```

Step.2: Implements the decision tree algorithm and returns accuracy, recall, precision and ROC AUC score.

Step.3: Implements logistic regression, Naive Bayes and Support Vector algorithm and returns accuracy, recall, precision and ROC AUC score.

```
dt_acc = accuracy_score(self.y_test, y_pred)
dt_precc = precision_score(self.y_test, y_pred)
dt_recall = recall_score(self.y_test, y_pred)
dt_auc = roc_auc_score(self.y_test, y_pred)
```

Step.4: Implements a neural network using the Keras library and returns accuracy and loss.

Step.5: Spit datasets into training and testing sets using one-hot encoding. And then Scaling is applied to both the datasets.

```
# Convert categorical variables in X to numerical values
using one-hot encoding
encoder = OneHotEncoder()
X_encoded = encoder.fit_transform(X)
# Split X_encoded and y into training and testing sets with
80:20 ratio
 X_train, X_test, y_train, y_test =
train_test_split(X_encoded, y, test_size=0.2)
```

Step.6: Accuracy, recall, precision, and ROC AUC score are computed using the Scikit-learn library functions. Additionally, k-fold cross-validation is performed.

```
 cvs = cross_val_score(estimator=model, X=self.x_train,
y=self.y_train, cv=10)
      print(cvs)
      return dt_acc,dt_recall,dt_precc,dt_auc
```

Step.7: Implements a neural network using the Keras library.It uses back-propagation to train the model. It consists of an input layer, one or more hidden layers, and an output layer.

```
From sklearn.neural_network import MLPClassifier
 # Create a MLP classifier with 2 hidden layers mlp =
MLPClassifier(hidden_layer_sizes=(100, 50),
max_iter=500)
# Train the model on training data mlp.fit(X_train, y_train)
 # Predict the class labels for test data y_pred =
mlp.predict(X_test)
```

IX. MATHEMATICAL MODEL

Let C denote the set of customers of a bank, and let t denote a discrete time period. For each customer c ∈ C, It define the following variables:

Churn(c,t) ∈ {0,1}

-Whether customer c has churned (left the bank) at time t.

- X(c,t) is a vector of predictor variables that describe customer c at time t, such as demographic information, account balances, transaction history, etc.

- **Y(c,t) ∈ {0,1}** is the target variable that here want to predict, which represents whether customer c will churn in the future (at some future time period t+Δt).

here assume that there exists a function f(X(c,t)) that maps the predictor variables to the target variable.

Y(c,t), i.e., f(X(c,t)) = P(Y(c,t)=1 | X(c,t))

This function can be learned from historical data using various data mining models such as logistic regression, decision trees, random forests, neural networks, etc[5].

Given this function f(X), this can use it to predict whether each customer will churn in the future by computing

f(X(c,t+Δt))

 DOI : 10.5281/zenodo.7856491

for each customer c at time t. It can then rank customers based on their predicted probability of churning and take appropriate actions to retain high-risk customers.

X. SYSTEM ARCHITECTURE

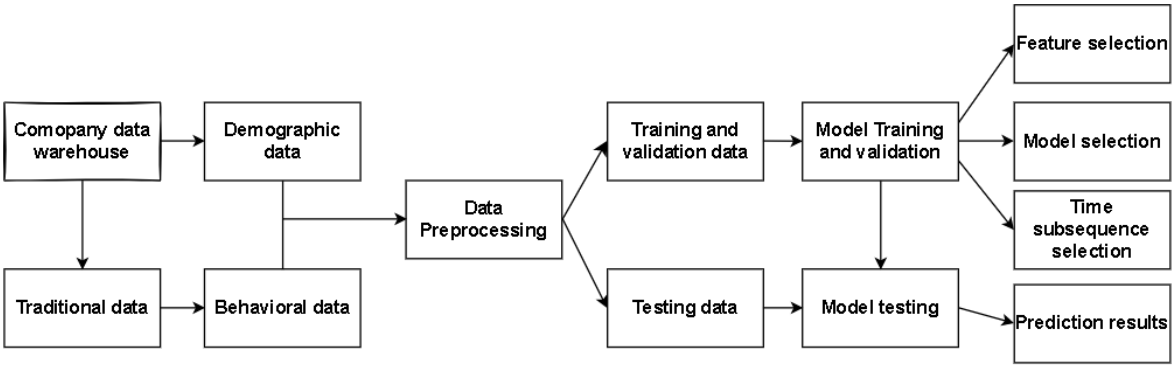

Fig.1: System Architecture

Step 1: Data of banks collected from various sources, which may be Traditional or Demographic data.

Step 2: Data is then prepossessed ,clean, sort. Then divided into Training data and Testing data.

Step 3: Training data then train through various training models, which results in selecting best model, features.

Step 4: Then testing data is test using that selected model to get prediction results.

XI. ADVANTAGES

- The model employed is decision tree, neural network, support vector machine (SVM), Naïve Bayes and regression logistic.
- All True Positive predictions are calculated in big amounts provided the funds have been effectively retained for not leaving the company.
- The value of hold able money is supposed to represent the profit obtained by the company if the model is executed. This is predicated on the premise for each customer that is recognized by churn and is genuinely churn after being followed up will still be a customer[7].

XII. DESIGN DETAILS

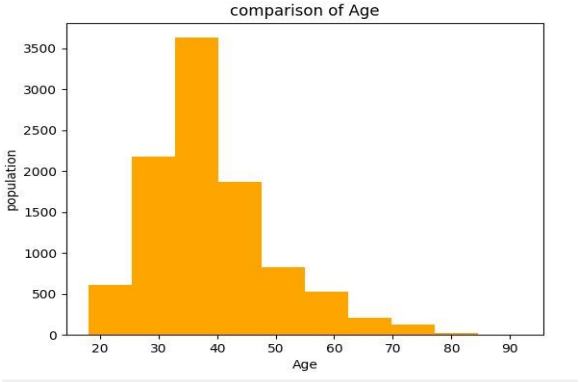

Fig 2:OutPut

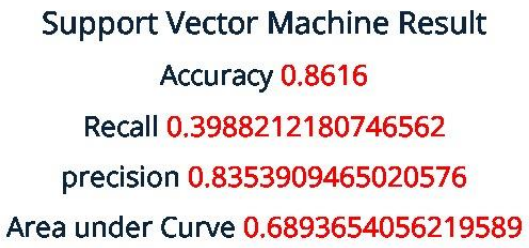

Fig 3:OutPut

- In fig 2 the population of churn customers represented according age.
- Fig 3 represents Accuracy, Recall, precision, AUC using SVM, which is more accurate algorithm for analysis.

XIII. CONCLUSION

Thus we have tried to implement the paper 'Ketut Gde Manik Karvana, Setiadi Yazid, Amril Syalim, and Petrus Mursanto(2019). Customer Churn Analysis and Prediction Using Data Mining Models in Banking Industry. IWBIS 2019 IEEE.' As The problem of customer turnover was observed to be becoming worse by the day, and prior efforts were examined to identify any gaps in the application of the remedy. Additionally, a collection of characteristics that were seen to influence churn were extracted using this method.

So as mentioned in above paper we try to implement Data mining model is used to predict customer churn in the

 DOI : 10.5281/zenodo.7856491

banking business. The number of samples of data sets used for training the data sets which greatly influences the results of modeling.

After training data sets with various algorithm it is concluded that the best model is the 50:50 SVM sampling model which gives approx 0.865 accuracy with a profit value of 456 billion with loss and benefit calculations.

Logistic Regression is also worth considering because it results in smaller losses. Accuracy values cannot be fully used as a reference for comparison if the distribution of data is very unbalanced.

REFERENCE

[1] Ketut Gde Manik Karvana, Setiadi Yazid, Amril Syalim, and Petrus Mursanto(2019). Customer Churn Analysis and Prediction Using Data Mining Models in Banking Industry. IWBIS 2019 IEEE.

[2] I. Ullah, B. Raza, A. K. Malik, M. Imran, S. U. Islam and S. W. Kim, "A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector," in IEEE Access, vol. 7, pp. 60134-60149, 2019, doi: 10.1109/ACCESS.2019.2914999.

[3] S. Agrawal, A. Das, A. Gaikwad and S. Dhage, "Customer Churn Prediction Modelling Based on Behavioural Patterns Analysis using Deep Learning," 2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE), Shah Alam, Malaysia, 2018, pp. 1-6, doi: 10.1109/ICSCEE.2018.8538420.

[4]Fenil Shah and Mrugendrasinh L. Rahevar, “Customer Churn Prediction Analysis,” November 2018 International Journal of Computer Applications 1829290:15-17.ResearchGate.

[5]Zoric, A. B. (2016). Predicting customer churn in banking industry using neural networks. Interdisciplinary Description of Complex Systems,

[6]Oyeniyi, A., & Adeyemo, A. (2015). Customer churn analysis in banking sector using data mining techniques. African Journal of Computing & ICT Vol 8, 165-174

[7]Chitra, K., & Subashini, B. (2011). Customer retention in banking sector using predictive data mining technique. International Conference on Information Technology. ICIT.

[8] Larose.D.T.(2006).Data mining methods and models. New Jersey:John Wiley & Sons, Inc.

[9]Gayatri Naik,”Tourist place reviews sentiment classification using machine, learing techniques, International Journal for Research in Engineering Application & Management,(IJREAM)ISSN:2454-9150,Volume 8 Issue 1.

 DOI : 10.5281/zenodo.7856491