CUSTOMER RETENTION USING

MACHINE LEARNING TECHNIQUES

Customer Churn And Retention

As organization market themselves to increase their market share against their competitors, they mainly focus on gaining new customers. However, they must also retain their existing customers. Since customer attrition can have a huge effect on customer growth and indirectly on the economy, even better will be is to focus on those customers with the highest value and the highest risk of closing their accounts.

For an Organization, retaining existing customers is usually easier and less expensive task as compared gaining new customers, since these customers already know and trust the organization, and also the organization already knows much about them.

The best possible way is to make the full use of the data we have on our customers,since customer attrition can have a huge effect on customer growth and even better will be to focus on those customers with the highest value and the highest risk of closing their accounts. With help of Machine Learning techniques what we can do is we can build some algorithms that will take the user past data as input and will predict whether there is chance that user might leave in future

The solution to keep the churn rate(rate at which customers stops doing business with an organization or entity) low is to identify in advance which customers(or say the group of customers) are going to leave and what are leading them to leave and then focusing on their retention programs(for retaining them).

Customer Segmentation Chart

Customers are divided into 4 quadrants based on their lifetime value and attrition risk.

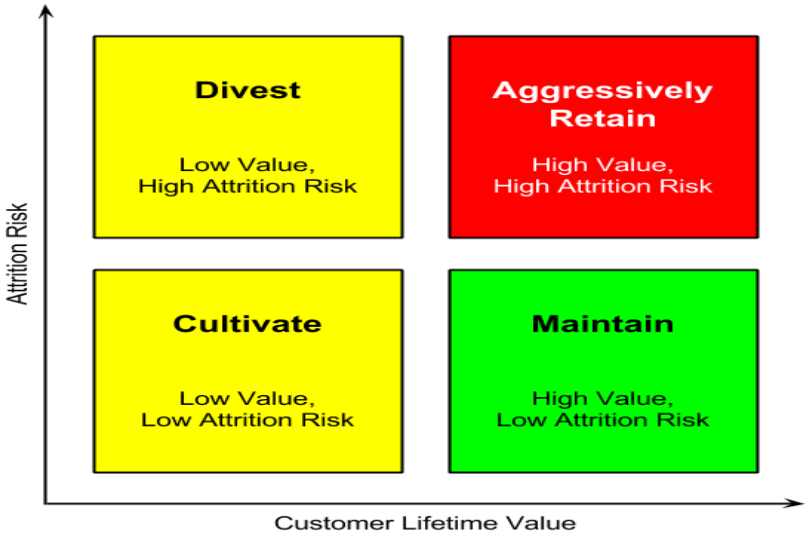
First step is to understand what kind of customers we have and what kind of attention they need.

1)Maintain-These are the customers with high lifetime value and low attrition risk. We can also say these are the our main customers and we should keep them satisfied. There chance of leaving is low and they are who add value to our organization.

2)Cultivate-These are the customers with low lifetime value and low attrition risk. They are also the one's which have low chance of leaving but aren't adding much value so we need to make them add high value(or say cultivate them).

3)Divest-These are the customers with low lifetime value and high attrition risk. We can also say that these are the customers which we take least care. They are the one's with high chance of leaving and low lifetime values.

4)Aggressively Retain-These are the customers with high lifetime value and high attrition risk.So we need to retain them any how as they are high value customers. They are the customers which are majorly focused so that they might not leave.



We need to focus differently on customers in different quadrants.

So the main concern is to focus upon the customers which have high attrition risk and high value and build a model on the existing data of our organization to predict whether a customer will close his or her account in future or not.

Data Preparation and Parameter Selection

Having data the next step is to Extract,Transform and Load the data.

1)The first and the foremost concern will be to analyze the customer data say what kind of data we have and what are the possible variables(or say features) we can have from our data. Analyzing the correlation and variance within our data,checking correlation between our input and output features. Finally taking the features which are correlated with our output variable and removing the redundant features. Identification of problem whether our problem is binary classification problem or multi class classification problem is also one of the main step.

2)Next step will be duplicating the data we have, using one for building the statistical model and other for using that statistical model to give the forecasts.

Building the statistical model we need to use the past data and taking output the customers that left the organization in past.

Using the statistical model for the future forecast we need to take the present data to forecast which customer will leave in future.

(Statistical model will be just only a mathematical equation which we need to figure out using different approaches so that our accuracy is high in predicting the customer mindset)

3)Another important step is to partition our data into training data,validation data and testing data.

Training data-To build each of the statistical model by training them on training data.

Validation data-Use validation data to check the performance of each of the models on validation set and picking the best one.

Test data-Test data is used to determine how well our best model from the bunch of models taken perform on our test set. We do this to make sure that everything is okay like in a case if the model performance on the test set is not what it was expected from its performance in validation set,we need to do some optimization.

Building The Statistical Model

The next step is the model building: As our problem identification has been done previous steps so now our main concern will be choosing the right mathematical model.

The best way for model selection is to try out a bunch of models and see how accurate are these models on our data, as we have a predefined knowledge of which models are better for certain task so what we can do is compare the precision and recall (or say the F1 Score) of the models and choose the one with highest F1 Score.

F1 Score = 2*(precision*recall)/(precision+recall)

The knowledge of the problem also help us in choosing the models like certain model perform good on binary task and others on Multi-Class Classification task.

The algorithms used and their performance:

1)Logistic Regression

The highest accuracy encountered using Logistic Regression was 0.8738.

2)Decision tree(Classification and regression trees)

Parameter used in CART is C P(complexity parameter).The model gave highest accuracy at C P = 0.0786.

3)Support vector machine

Parameters needed to be tuned in SVM are C and sigma. C and sigma used for the model are 1 and 0.0629 at highest accuracy.

4)K-nearest Neighbor

In KNN the parameter that need to be tuned is k and results show that model gave highest accuracy at k = 7.

5)Ensemble Learning Techniques:

Ada Boost

In Ada Boost nIter is the parameter that needed to be determined. The model gave highest accuracy at nIter=100.

Stochastic Gradient Descent

Initially the number of trees were taken between 5000 to 10000 and the model was tuned to calculate the number of trees that achieve the best accuracy and at number of trees equal to 6000 highest accuracy was encountered.

Random Forest

Initially 500 trees were taken but the error rate results indicate that after 100 decision trees there was no significant reduction in error rate and mtry was 7.

6)Neural Networks

Multi-layer Perceptron was used with 13 input neuron and 5 neuron in hidden layer and 2 output neuron was taken. At learning rate of 0.01 maximum accuracy was encountered.

All of these model works better than other for certain set of condition and certain type of data but when these models were tested upon customer data of a telecommunication company we got different accuracy corresponding to different models.

The data contains customers statistical data and includes 17 features related to customers service:

It was a binary classification problem.

14% of the observations have the target variable yes.

86% of the observations have the target variable No.

Models with their Accuracy

Model	Min	Max
Random Forest	0.9311	0.9639
Ada Boost	0.9341	0.9649
Multi Layer Perceptron	0.9332	0.9440
Stochastic Gradient Descent	0.8900	0.9439
Support Vector Machine	0.9041	0.9401
K-Nearest Neighbor	0.8738	0.9159
CART	0.8528	0.9039
Naive Bayes	0.8648	0.8828
Logistic Regression	0.8443	0.8738
Linear Discriminant Analysis	0.8328	0.8678

Based upon the model we used, we got highest accuracy using Random Forest.

Using Random forest for such task can be a great choice but as the data format from one problem to another problem changes the performance of models also get affected so better option will be to try the problem with the top 5 algorithms from table and compare them. Trying various advanced neural network architectures using transfer learning techniques can also help in getting higher accuracy.

Providing the Model Output to the Marketing Team

After getting the predictions from the best chosen model the churners along with their data will be passed on to the sales and the marketing team, so that they can retain the customer using some retention strategies and also gain understanding as what are the factors that lead to customer leaving.

References:

1) Enhancing Customer Retention Through Data Mining Techniques
2) Machine-Learning Techniques for Customer Retention: A Comparative Study