Predicting customers at risk with machine learning

David Gojo
Jožef Stefan International Postgraduate School,
Jamova 39, 1000 Ljubljana, Slovenia
david.gojo@ijs.si

Darko Dujič
Ceneje d.o.o.,
Jožef Stefan International Postgraduate School,
Štukljeva cesta 40, 1000 Ljubljana, Slovenia
darko.dujic@ceneje.si

ABSTRACT

Today's market landscape is becoming increasingly competitive as more advanced methods are used to understand customer's behavior. One of key techniques are churn mitigation tactics which are aimed at understanding which customers are at risk to leave the service provider and how to prevent this departure. This paper presents analyzes accounts renewal rates and uses easily applicable models to predict which accounts will be decreasing spend at the time when they are due to renew their existing contract based on number of attributes. Key questions it tries to explore is if customer behavioral or customer characteristic data (or combination of both) is better at predicting accounts that will renew at lower than renewal target amount (churn rate).

Categories and Subject Descriptors

F.2.1 **[Numerical Algorithms and Problems]:** Data mining, Structured prediction

General Terms

Algorithms, Management, Measurement, Documentation, Performance

Keywords

Data Mining, Analysis, Churn prediction.

1. INTRODUCTION

The main issue of business is how to make educated decision with support of analysis that dissect complex decisions on addressable problems using measurements and algorithms. Where there are many disciplines are researching methodological and operational aspects of decision making, at the main level, we distinguish between decision sciences and decision systems [1]. With increasing number of companies trying to use machine learning to assist in their decision-making process we examined how decision science can be supplemented by applying machine learning models to the company's customer data. We partnered with the medium sized B2B business operating in Europe and Africa with the aim to help them better understand their 'customers at risk' segment of clients.

To this end we developed two easily applicable performance algorithms designed to highlight customers at risk and company can address to mitigate their risk of leaving as clients.

The paper has the following structure: in section 2 we are presenting related work to the area recorded historically. Next, data acquisition is explained in section 3 followed by results acquired from the tested algorithms in section 4. We then conclude our observations in section 5.

2. RELATED WORK

Improvements in tracking technology have enabled data driven industries to analyze data and create insights previously unavailable to the business. Data mining techniques have evolved to now support the prediction of behavior of customers such risk of leaving due to the attributes that are trackable [2]. The use of data mining methods has been widely advocated as machine learning algorithms, such as random-forest approaches have several advantages over traditional explanatory statistical modeling [3].

Lack of predefined hypothesis makes algorithms excel in these tasks as it is making it less likely to overlook predictor variables or potential interactions that would otherwise be labelled unexpected [4]. Models are often labelled as business intelligence models aimed at finding customers that are about to switch to competitors or leave the business [5].

Key classifications are observed in work related to churn that we will use in our data set for review [6]:

- Behavioral data - will consist of attributes that we have historically observe that play a role in whether the account will renew or not: product utilization, activity levels of the account, number of successful actions in the account and number of upsells done ahead of renewal.
- Characteristic attributes - will consist of size of the account in terms of spend, number of members in the company, number of active users of the products in the company, payment method and how they renew the contract, geography and what level of support the product is given (number of sales visits and interactions with the customer).

3. DATA ACQUISITION

3.1 Data understanding

Working with the customer we have arranged a set of interviews with the leadership to better understand their business and challenges they are experiencing together with ambitions they have in the business. After the interview rounds we focused on reviewing 2 hypotheses flagged in the examination process:

- What is driving churn numbers: behavior of the customers or better structure of the base?

- Does acquisition of new accounts represent a risk in churn number with historic observation of accounts renewing lower / not renewing in their first-year renewal?

3.2 Data pre-processing

The data we used derives from company's internal customer bookings and customer databases we consolidated. As customers are on yearly renewals we have taken the renewal and the data on the account before the renewal as the key building block for analysis.

3.3 Feature engineering

We enriched the data using SQL joins on the customer numbers to include key characteristics of accounts, tenure of the client, products utilization information, behavior of the customer before the renewal and their usage of the product.

In terms of regional split of the market the dataset consists of 4 key geo and segment regions in Europe and Africa:

- Medium-business segment
- UK & Ireland market
- Europe Enterprise segment
- Eastern Europe, Middle-East and Africa

Through feature engineering and reviewing descriptive statistics key attributes we nominalized are 11.

For the machine learning purposes for the calls we have selected 3 possible outcomes related to the outcome of customer spend after it's renewal:

- Customer_Renew (Not renew, Partial renew, Full renew)

3.4 Data Set Statistics

We selected bookings period from 2016 to end of 2017 including 23,043 instances in above selected renewal of 12,872 accounts. The attributes that were nominalized are listed below:

- (nom) Has main product – has product 1
- (nom) Has_assisting_product – has product 2
- (nom) Has_media_product – product 3
- (nom) Account_potential – size and potential of the account
- (nom) Is_Auto_Renew – auto renewal option enabled
- (nom) First_renewal – is the client renewing first time
- (nom) Upsold_Before_renewal – was there an upsell before
- (nom) JS_Utilization – utilization of product 2 - indicator
- (nom) Score_Engagement – engagement of the recruiter
- (nom) LRI_Score – savviness of the user of the product

4. RESULTS

4.1 Brief description of the methods used

Where multiple algorithms were used during the testing due to important feature that the result needed to have at least one interpretable model, so we went in the direction of nominalizing attributes and decided to use J-48 model and Random forest classifier on the data set.

J48. Decision trees C4.5 (J48 in Weka) algorithm: deals with continuous attributes as observed in the related work.

Where the method is classification-only the main machine learning method applied is J48 pruned tree or WEKA-J48 machine learning method. Tree tries to partition the data set into subsets by evaluating the normalized information gain from choosing a descriptor for splitting the data. The training process stops when the resulting nodes contain instances of single classes or if no descriptor can be found that would result to the information gain.

Random Forest. We assume that the user knows about the construction of single classification trees. Random Forests grows many classification trees. To classify a new object from an input vector, put the input vector down each of the trees in the forest. Each tree gives a classification, and we say the tree "votes" for that class. The forest chooses the classification having the most votes (over all the trees in the forest) [7]. Both methods were applied to imported dataset numerous times with continuous testing of parameters to improve performance.

4.2 Application of J48

Working with Weka on the dataset of the customer we tried to launch the model to tune the parameters. Key modifications:

- "10-fold cross validation" used to improve accuracy
- Minimum number of objects moved to 100

As Figure 2 shows this reduced the number of leaves to 16 which was something comprehendible enough.

Summary of results below:

```
=== Summary ===

Correctly Classified Instances       16789               72.8626 %
Incorrectly Classified Instances      6253               27.1374 %
Kappa statistic                          0.249
Mean absolute error                      0.2759
Root mean squared error                  0.3716
Relative absolute error                 88.6325 %
Root relative squared error             94.189  %
Total Number of Instances            23042
```

Figure 1: J-48 model error estimates

4.3 Application of Random forest

We ran several tests on Random forest vs Random trees. When tuning parameters Random tree tended to not respond well to pruning so Random forest was a preferred option. Like J48, application with key modifications was focused on validation and additionally on setting maximum depth of the random forest:

- "10-fold cross validation"
- Max. depth set at 3

Summary of results below:

```
=== Summary ===

Correctly Classified Instances       16635               72.1943 %
Incorrectly Classified Instances      6407               27.8057 %
Kappa statistic                          0.1852
Mean absolute error                      0.28
Root mean squared error                  0.3685
Relative absolute error                 89.9598 %
Root relative squared error             93.401  %
Total Number of Instances            23042
```

Figure 2: Random forest model error estimates

4.4 Comparisons of models

Overall the J48 model has surprisingly 0.7pp points higher Classification Accuracy than the Random forest model.

Validation Measures	J48	Random Forest
Classification Accuracy	72.9%	72.2%
Mean absolute error	0.276	0.280

Table 1. Baseline benchmark validation measures

Key observation analyzing the data was that neither model was predicting any partially churned accounts after we forced their trees to be pruned.

J48 predictions:

a	b	c		<-- classified as
0	2745	285	|	a = PARTIAL_RENEW
0	1528	789	|	b = FULL_RENEW
0	2434	1504	|	c = NOT_RENEWED

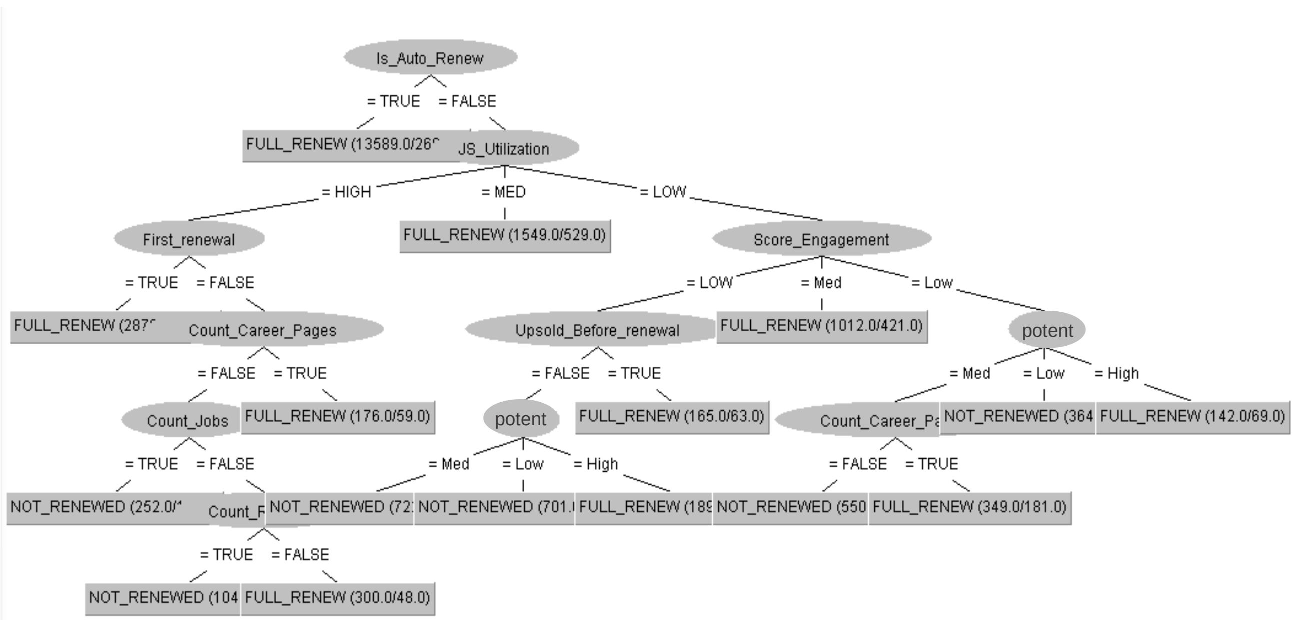

Figure 3: The J48 decision tree

Random forest predictions:

a	b	c		<-- classified as
0	2857	173	|	a = PARTIAL_RENEW
0	15591	483	|	b = FULL_RENEW
0	2894	1044	|	c = NOT_RENEWED

Even though Random forest has a lower classification accuracy J48 offers significantly higher interpretability with tree pruning offering valuable insights, short description below and discussed in evaluation of models.

J48 provided a significantly better interpretability and classification accuracy than the Random forest or any test on the Random tree model. Some additional tests were done on Naïve Bayes model and J48 was superior in the results. Key area it accelerated was in predicting accounts that will not renew. Where the precision is just above 38% this is almost double comparing to Random forest model.

3 key takeaways observed that the company found the most insightful were:

- One of the new features designed by the product team that encouraged the auto-renew of their clients played the most important at predicting the renewal rate
- Customer behavior is a better signal for probability of renewal vs general account characteristics
- Account potential which is the predictor of account potential and size plays the role only after product utilization and engagement of the account with our products

5. CONCLUSION AND FUTURE WORK

For the acceleration of performance, the decision tree is of paramount importance and value. Insight that Auto renew as a feature is one of the key predictors if the account will renew fully is truly remarkable based on the simplicity of the models and how easily applicable they are.

Applications of this models will be of great foundation for driving the discussion on different account features and metrics. This is especially true as it is tackling one of the key challenges observed in hypothesis as in how important 'account potential' is for the account ahead of the renewal.

Observing the attributes added into the analysis scope and optimizing them for the J48 we were able to get valuable insight which account characteristics vs account behaviors ahead of the renewal are the best predictors for the account to renew at the full amount.

6. REFERENCES

[1] M. Bohanec, Decision Making: A Computer-Science and Information-Technology Viewpoint, vol. 7, 2009, pp. 22-37.

[2] A. Rodan, A. Fayyoumi, H. Faris, J. Alsakran and O. Al-Kadi, "Negative correlation learning for customer churn prediction: a comparison study.," *TheScientificWorldJournal,* vol. 2015, p. 473283, 23 3 2015.

[3] A. K. Waljee, P. D. R. Higgins and A. G. Singal, "A Primer on Predictive Models," *Clinical and Translational Gastroenterology,* vol. 5, no. 1, pp. e44-e44, 2 1 2014.

[4] Y. Zhao, B. Li, X. Li, W. Liu and S. Ren, "Customer Churn Prediction Using Improved One-Class Support Vector Machine," Springer, Berlin, Heidelberg, 2005, pp. 300-306.

[5] M. Óskarsdóttir, B. Baesens and J. Vanthienen, "Profit-Based Model Selection for Customer Retention Using Individual Customer Lifetime Values," *Big Data,* vol. 6, no. 1, pp. 53-65, 3 2018.

[6] S. Kim, D. Choi, E. Lee and W. Rhee, "Churn prediction of mobile and online casual games using play log data," *PLOS ONE,* vol. 12, no. 7, p. e0180735, 5 7 2017.

[7] J. Hadden, A. Tiwari, R. Roy and D. Ruta, "Computer assisted customer churn management: State-of-the-art and future trends," *Computers & Operations Research,* vol. 34, no. 10, pp. 2902-2917, 10 2007.

[8] A. K. Meher, J. Wilson and R. Prashanth, "Towards a Large Scale Practical Churn Model for Prepaid Mobile Markets," Springer, Cham, 2017, pp. 93-106.

[9] M. Ballings, D. Van den Poel and E. Verhagen, "Improving Customer Churn Prediction by Data Augmentation Using Pictorial Stimulus-Choice Data," Springer, Berlin, Heidelberg, 2012, pp. 217-226.