Telco customer churn analysis: measuring the effect of different contracts

Paulo Pinheiro[1][0000-0002-8912-2244] and Luís Cavique[2][0000-0002-5590-1493]

[1] Universidade Aberta and Cedis, Lisboa, Portugal
[2] Universidade Aberta, Lisboa, Portugal
ppinheiro@cedis.pt, luis.cavique@uab.pt

Abstract. Customer retention is nowadays a challenge that requires concrete and personalized actions. Traditional data mining studies focused on predictive analytics, neglecting the business domain. This work aims to present an actionable knowledge discovery based on specific, actionable attributes and measuring of their effects. It is common to use matching, and propensity score approaches in healthcare to evaluate causality. After performing matching using the actionable attributes in this analysis, the causal effect is quantified. This work concludes that the difference between having a yearly contract versus having a monthly contract affects the churn of around 34%.

Keywords: customer churn, actionable attributes, causality, propensity score matching

1 Introduction

Churn rate refers to the percentage of customers who leave a supplier during a specific period. Various reasons can be associated, such as customer dissatisfaction, better competition offers, or customer lifecycle.

To get actionable knowledge from data that allows identifying customers at risk of moving out and those who keep being loyal to the service is a relevant issue. In this context, the concept of actionable knowledge can be successfully applied since the goal is not focused on the predictive algorithms but on solving business problems [1].

Traditional data mining focuses essentiality on predictive mining. However, the information by itself does not benefit the end-user. What becomes more critical to organizations is to mine patterns in order to create actionable knowledge. Actionable knowledge is supported by actionable attributes, that is, attributes that can be manipulated and allow operational changes. Actionable attributes must make explicit the causal processes, so they are the key in the interventions. In this customer churn study, the actionable attribute is the type of contract.

The objective of the work is to find a method to discover actionable attributes and measure the causal effect of those attributes to define their impact on the business.

The contribution of this work is the application of causal effect methods in the specification of actionable knowledge. Biostatistics techniques like propensity score matching are used [2].

The paper is organized into five additional sections. In Section 2, related work is presented. Section 3 introduces the causal effect approach. Section 4 details the Telco case study. Section 5 discusses the business impact. Finally, in Section 6, conclusions are drawn.

2 The customer churn problem

Churn management is a top priority for most businesses because it is directly tied to firm profitability and value [3]. According to Surujlal and Dhurup [4], there are several benefits associated with customer retention: the cost of acquiring a customer only takes place at the beginning of the relationship and therefore the longer the relationship, the lower the cost; the likelihood of long-term customers switching is low, tends to be less price sensitive and more likely to give referrals; they are also more likely to purchase ancillary products and supplements that have higher margins; and because of their knowledge of the organization, the cost of serving them is lower.

Mahajan *et al.* [5] in a literature review on Churn in telecommunications, points out a framework for the types of service abandonment, the support of studies on this phenomenon in obtained data, and the use of data mining as a way to create predictive models. They refer that in some sectors, given the nature of how their services are provided, they suffer from high dropout rates, particularly in the telecommunications sector. In this industry, customer churn can happen voluntarily or involuntarily. Voluntary withdrawal happens when a customer decides to leave a service provider and join another, which can happen because something happened in customer's life (change of financial conditions, change of residence) or in a deliberate way, motivated by the desire to obtain better technology, better service, better price, etc. The involuntary withdrawal occurs without the customer's will when, for example, he fails to pay.

Researchers have been using a large amount of data at their disposal to study and seek to mitigate the churn problem, given the large number of subscribers they have, rapidly renewable technologies, database applications and other value-added services. These databases, in addition to the large number of records relating to customers, also have a variety of attributes about the customer and the various services subscribed to by them, like customer demographic data (age, gender, tenure, income), internal customer plan (plan type, billing agency, billing disputes, number of weekly calls, national and international call billing), call details (call duration, call type), and others.

Data mining is often used in Churn Analysis applications in the telecommunications industry to find correlations or patterns between attributes of large databases. It is used to predict the behavior of customers who are most likely to quit the existing provider's services and join new service provider. Understanding the current and past trends, behavior, and planning for the future is essential in business. Hence, data mining applications play an essential role in decision making and providing prediction on future estimates.

However, traditional data mining studies concentrated primarily on predictive mining, describing the cause and effect scenario is described. However, this information alone is not sufficient as it does not benefit the final user. What becomes more exciting and critical to organizations is to mine patterns to create actionable knowledge [1].

Actionability should be a criterion that can measure the utility of the mined patterns. Knowledge is considered actionable if users can take direct actions based on such knowledge to their advantage. Among the most critical and distinctive actionable knowledge are actionable behavioral rules that can directly and explicitly suggest specific actions to influence (retain and encourage) the behavior of customers [6].

3 The causal effects

When studying causality, it is common to hear 'correlation does not imply causation' or 'correlation is not causation.' Spurious correlations are a strong association of two or more variables due to a pure coincidence or a presence of a third variable, named the confounder. A set of interesting examples can be shown in Spurious Correlations [7].

3.1 Prediction vs. causality

In data science, two tasks must be distinguished: prediction and explanation. In prediction, two variables are used, the independent variable X and the dependent variable Y. The original data is divided into the training and testing data sets to find the function f (X, Y), where X is a covariate and Y is the outcome.

A new variable type should be included in the causality task; the intervention/treatment T. In this task, the outcome Y of treatment T is the subject of the study. For this purpose, test and control datasets are used for treatment accomplished T=1 and not accomplished T=0. In analogy with f (X, Y), the explanatory function uses three variables, f (T, X, Y). This dichotomy can be found in the ladder of causation of Pearl and Mackenzie [8] and the work of Hernán *et al.* [9].

Fisher [10] introduced the experimental design, which uses randomized trials with an experiment sample, and a control sample, also known as the A/B test. This method is a standard in science. However, in social science and healthcare A/B test is not considered adequate, not only because of the high number of particular cases but also because it is considered unfeasible or unethical. Moreover, in many economic activities, the historical data does not present the two referred samples. Instead of sampling and experimental design, the focus is on quasi-experimental design, particularly in observational studies.

3.2 Direct Acyclic Graphs

The Direct Acyclic Graphs (DAG), is a handy tool in causal representation; they describe the causal assumptions of each study [11]. The nodes correspond to the variables (treatment, covariates, or outcome) and the arrows are the eventual association between the nodes. The data description and the DAG should come before the models [12].

Using the same variables with arrows with distinct directions represents different DAG. Figure 1 (on the left) shows the variable X as a confounder of T and Y. Figure 1 (on the right) variable X is a mediator between T and Y.

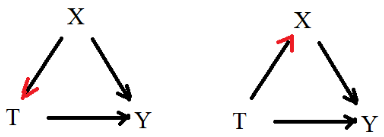

Fig. 1. DAGs: X is a confounder (left), and X is a mediator (right)

3.3 Paradoxes and adjustment

In the following example of confounder variable X is the gender. When dealing with stratified treatments by gender, the treatment can be good for people but harmful for males and females. This contradiction is called the Simpson's paradox and can be solved using the adjustment formula [11], as follows:

$$P(Y = y|do(T = t)) = \sum_{|G|} P(Y = y|T = t, G = g). P(G = g) \quad (1)$$

When there is more than one confounder, standardization of two or more variables has many possible combinations. So, other methods like propensity score matching are used.

3.4 Matching and propensity scores

In pairwise matching, the individuals are matched on an individual basis in two groups. In a treatment–control study, each treated is matched individually to control who has similar risk factors (X->Y). In the matching process, the effects of any potential confounding variables X will be removed without the need to adjust for these variables in the analysis.

Based on a vector of observed covariates, the propensity score is the conditional probability of assignment to a particular treatment [13]. The simplest method of estimation is using a logit model using data collected from both groups. The propensity score matching algorithm retrieves the matching of treatment and control individuals who have similar propensity scores. Finally, the distribution of propensity scores should be similar across the groups of treatment and control. The disadvantage of this approach is that some individuals are excluded from the analysis if a suitably matched pair cannot be identified.

3.5 Measuring the impact of the treatment

The treatment effect can be given by the linear regression:

$$Y = a + b.T + c.X + e \tag{2}$$

where *e* represents the error and the slope *b* measures the causal effect [14].

For the special case of the bivariate regression:

$$Y = a + b.T + e \tag{3}$$

the conditional expectation Y given the treatment T, takes two values, as follows:

$$E(Y|T=0) = a \tag{4}$$

$$E(Y|T=1) = a + b \tag{5}$$

and then

$$b = E(Y|T=1) - E(Y|T=0) \tag{6}$$

is the difference in expected Y with treatment T.

Using this notation:

$$E(Y|T) = E(Y|T=0) + [E(Y|T=1) - E(Y|T=0)]. T = a + b. T \tag{7}$$

So, E(Y|T) is a linear function of T, with slope *b* and intercept *a*. The regression slope measures the difference in expected Y with treatment T switched on and off.

4 Telco case study

This case study presents the Telco Customer Churn public dataset [15] that contains information on eighteen covariates potentially related to both the outcomes of interest (churn or not churn). Our goal is to find any actionable attributes, in which we can intervene to avoid dropouts and measure their causal effects and impact on the business.

4.1 Dataset description

In this dataset, each row represents a customer and each column contains customer's attributes. Those attributes can be grouped in demographic info about customers, like gender, age range, and if they have partners and dependents; attributes that describe customer's account information, like how long they have been a customer, contract, payment method, paperless billing, monthly charges and total charges; and attributes that present the services that each customer has signed up for, like phone, multiple lines, internet, online security, online backup, device protection, tech support and streaming TV and movies. There is also an attribute, Churn, which indicates whether or not the customer has abandoned services in the last month.

The dataset contains 7043 unique customers, of which 5174 have dropped out of services, and 1869 continue to use the services, which show that churn distribution is unbalanced.

Imbens and Rubin [16] mentioned that it is not always feasible to include all covariates in a model for the propensity score. Moreover, it may not be sufficient to include only linearly for some of the important covariates. We may wish to include functions like logarithms, higher-order terms, quadratic terms, or interactions between the primary covariates. The number of covariates was reduced, choosing the most relevant for the business. In order to facilitate the presentation of this work, the covariates were binarized. Table 1 shows the resulting attributes and the transformation performed on their values.

Table 1. Used attributes and value transformations

Attribute name	Binarized	
	0	1
Gender	Female	Male
Partner	No	Yes
Dependents	No	Yes
PhoneService	No	Yes
Contract	Month-to-month	One Year, Two Year
Churn	No	Yes

In this case study, the actionable attribute is the *Contract* attribute. For monthly contracts, the binarized value is 0, and 1 denotes loyalty for one or two years.

4.2 Direct Acyclic Graph

Observing the dataset concerning the Contract and Churn attributes, we conclude that 55% of the contracts are monthly and in this customer segment there is 42.71% of dropouts. Considering that the Contract attribute is actionable, it can target an intervention to make customers switch to long-term contracts to reduce churn.

An intervention can be planned to improve the recovery rate by increasing the duration of the customers' contract to one or two years. Covariate T is the intervention to make the customer increase the duration of the contract, the Y as the outcome, meaning that the customer may or may not cancel the service, and X the set of confounding factors that may be interfering with the outcome. The DAG obtained is illustrated in Figure 2.

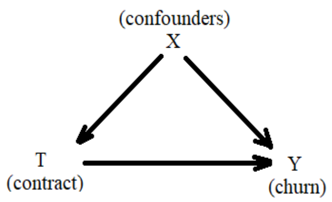

Fig. 2. DAG for Contract treatment

If there were only one confounder, we could use the formula given in (1) to obtain what is known as the Average Causal Effect (ACE) [11]. However, X is a set of attributes that can take different combinations, so we will use propensity score matching, as described in the next section.

4.3 MatchIt and Propensity Score

Austin [17] states that the propensity score is a balancing score: conditional on the propensity score, the distribution of observed baseline covariates will be similar between treated and untreated subjects. He also describes several methods for propensity score. The method use a binary treatment on an outcome while controlling for measured pre-treatment confounding variables [18]. In this work, after applying the propensity score matching, the causal effects are estimated.

To do this, we used system R with library MatchIt, Nonparametric Preprocessing for Parametric Causal Inference, Ver. 4.3.0, as described in R documentation which enables parametric models for causal inference to work better by selecting well-matched subsets of the original treated and control groups.

To perform matching, we used 1:1 nearest neighbor matching on the propensity score, which is appropriate for estimating the Average Treatment Effect in the treated (ATT), the average treatment effect for the units. We also specify *glm* (generalized linear model) as *distance* parameter as the propensity score method. The results obtained are shown in Table 2 and 3.

Table 2. Sample Sizes

Sample Sizes	Control	Treated
All	3875	3168
Matched	3168	3168
Unmatched	707	0
Discarded	0	0

Table 3. Balance for All and Matched data

Attribute name	All data			Matched data			
	Std.Mean Diff	eCDF Mean	eCDF Max	Std.Mean Diff	eCDF Mean	eCDF Max	Std.Pair Dist.
Distance	0.6156	0.1946	0.2831	0.4549	0.1362	0.2030	0.4549
Gender	0.0068	0.0034	0.0034	-0.1376	0.0688	0.0688	0.9041
Partner	0.5872	0.2821	0.2821	0.4217	0.2027	0.2027	0.4217
Dependents	0.4327	0.2134	0.2134	0.3406	0.1679	0.1679	0.6056
PhoneService	0.0015	0.0004	0.0004	0.0748	0.0221	0.0221	0.6155

As we can see in Table 2, each of the 3168 treated units was paired with an available control unit with the closest propensity score. The remaining 707 control units are left unmatched and excluded from further analysis.

In Table 3 we can see the values obtained with all the data (columns 2 to 4) and after the matching (columns 5 to 7). More outstanding balances are found when comparing standardized mean differences (*Std. Mean Diff.*) and empirical cumulative density function (*eCDF*) statistics for all data and matched data, as the latter approach zero. The last column, *Std.Pair Dist*. presents the absolute mean difference between the pairs for each covariate so that a better balance is obtained when the value it presents is smaller.

Figure 3 and 4 also show a matching distribution and histograms of the propensity scores for matched control and treated units.

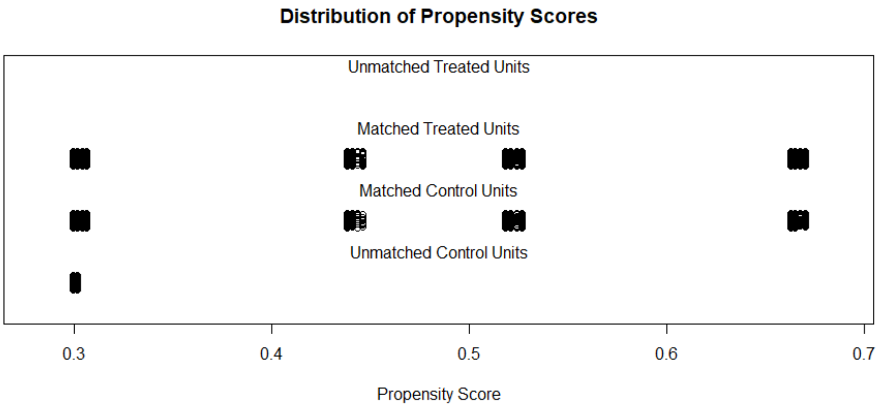

Fig. 3. Distribution of Propensity Scores for Contract treatment

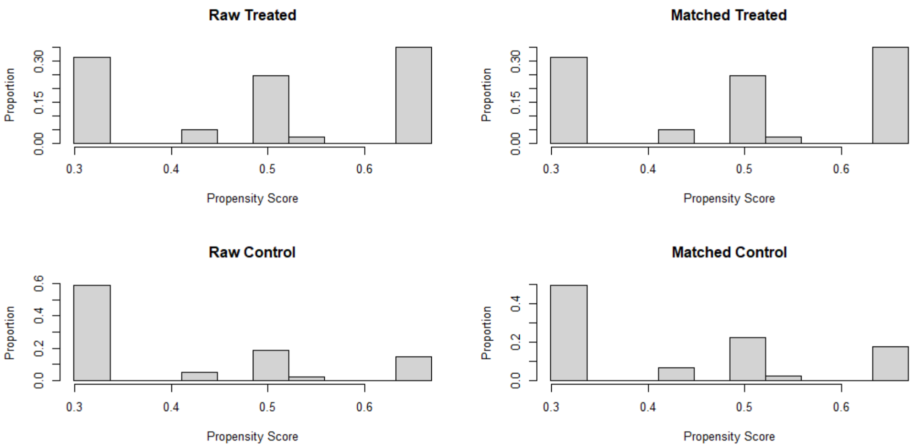

Fig. 4. Histograms of Propensity Scores

4.4 Causal effect calculations

To estimate the average effect of treatment on “treated” customers, a regression function in R (*lm*) was used, obtaining the Contract coefficient, which is an estimated value for ATT = -0.33757019. Thus, the difference of customers with monthly contracts compared with the customers with contracts of 1 to 2 years has an effect on the churn of around 34%.

Knowing that 57% of clients with month-to-month contracts drop out, the previous result concludes that it is possible to recover almost 34% of them. Also, knowing that the average monthly charge of these customers is 61.46, the recovery of these customers allows recovering up to 46154.92 in the total monthly amount. This value also represents a maximum estimate for the amount spent on recovering these customers without incurring losses.

5 Conclusion

Telco's churn is around 26%, revealing the importance of customer retention interventions that require concrete and personalized actions. Most data mining techniques are focused on predictive analytics but neglect the business domain.

In this work, the focus in creating actionable knowledge in order to develop a retention program. Actionable attributes support actionable knowledge, that is, attributes that can be manipulated and allow operational changes. In this case study, the actionable attribute is the type of contract since customers with annual or biannual contracts tend to be more loyal than customers with a monthly contract.

The operational change encourages customers with a monthly contract to invest in an annual contract in exchange for benefits. In order to assess the possible benefits, the causal effect of the type of contract has to be evaluated.

The techniques applied to identify the causal effects are often used in biostatistics, like propensity score matching. The contribution of this work is the application of those methods in the specification of actionable knowledge.

The causal effect retrieved by the regression model is around 34%, which allows evaluating the budget for the encouragement campaign.

This approach has some limitations since the customers were not randomly chosen, incurring in selection bias. We pretend to include techniques as instrumental variables to avoid these constraints in future work.

References

[1] L. Cao, “Domain-driven data mining: Challenges and prospects,” *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 6, pp. 755–769, 2010.

[2] M. A. Hernán and J. Robins, *Causal Inference*. CRC Press, 2011.

[3] E. Ascarza, “Retention futility: Targeting high-risk customers might be ineffective,” *J. Mark. Res.*, vol. 55, no. 1, pp. 80–98, 2018.

[4] J. Surujlal and M. Dhurup, “Establishing and maintaining customer relationships in commercial health and fitness centres in South Africa,” *Int. J. Trade, Econ. Financ.*, vol. 3, no. 1, pp. 14–18, 2012.

[5] V. Mahajan, R. Misra, and R. Mahajan, “Review of Data Mining Techniques for Churn Prediction in Telecom,” *J. Inf. Organ. Res.*, vol. 39, no. 2, pp. 183–197, 2015.

[6] P. Su, D. Zhu, and D. Zeng, “A New Approach for Resolving Conflicts in Actionable Behavioral Rules,” *Sci. World J.*, vol. 2014, 2014.

[7] T. Vigen, “Spurious Correlation,” 2015. [Online]. Available: https://www.tylervigen.com/spurious-correlations. [Accessed: 09-Nov-2021].

[8] J. Pearl and D. Mackenzie, *The book of why: the new science of cause and effect*. Basic Books, 2018.

[9] M. A. Hernán, J. Hsu, and B. Healy, “A Second Chance to Get Causal Inference Right: A Classification of Data Science Tasks,” *Chance*, vol. 32, no. 1, 2019.

[10] R. A. Fisher, *The design of experiments*, 8th ed. Hafner Publishing Company, New York, 1966.

[11] J. Pearl and M. Glymour, *Causal Inference in Statistics: A Primer*. Wiley, 2016.

[12] M. A. Hernán, “Causal Diagrams: Draw Your Assumptions Before Your Conclusions,” 2017. [Online]. Available: https://www.edx.org/course/causal-diagrams-draw-your-assumptions-before-your. [Accessed: 12-Nov-2021].

[13] P. R. Rosenbaum and D. B. Rubin, “The Central Role of the Propensity Score in Observational Studies for Causal Effects,” *Biometrika*, vol. 70, no. 1, pp. 41–55, 1983.

[14] J. D. Angrist and J.-S. Pischke, *Mastering Metrics: the path from cause to effect*. Princeton University Press, 2015.

[15] BlastChar, “Telco Customer Churn,” 2018. [Online]. Available: https://www.kaggle.com/blastchar/telco-customer-churn. [Accessed: 01-Nov-2021].

[16] G. W. Imbens and D. B. R. Rubin, *Causal Inference for Statistics, Social, and Biomedical Sciences: an introduction*. Cambridge, 2015.

[17] P. C. Austin, “An introduction to propensity score methods for reducing the effects of confounding in observational studies,” *Multivariate Behav. Res.*, vol. 46, no. 3, pp. 399–424, 2011.

[18] N. Greifer, “MatchIt : Getting Started,” 2021. [Online]. Available: https://cran.r-project.org/web/packages/MatchIt/vignettes/MatchIt.html. [Accessed: 09-Nov-2021].