Telco Customer Churn Recommendation

Fall 2 Blue 17

Feliciti Milne
Josh Hull
Noah Johnson
Jiwon Lee
Patrick O'Keefe

October 17th 2022

Table of Contents

Telco Customer Churn Recommendation

Overview

Telco has asked our team to build a decision tree model to predict churn in order to compare it with their in-house model. In response, our team has built a classification tree model and a recursive decision tree model using Telco's customer churn data. Both models identified similar criteria for splitting the data that resulted in only slightly different outcomes. Our team therefore recommends utilizing the classification tree model based on the simplicity of the model and the resulting clarity of business focus. Using these models and the resulting customer segments, our team has identified fiber optic customers on month-to-month contracts as the highest priority customer segment. This group made up 31% of the customers in the training data, however they accounted for over 41% of monthly revenue. In addition, this group is predicted to churn at a 55% probability.

Methodology

Data Pre-Processing

We received churn data from Telco that contained 7043 observations and 21 variables. There were 11 rows with missing values in the Total Charges column, which corresponded to new customers with a tenure of 0. Therefore, we opted to input these missing values with zeros since the customers were new and had not accumulated any charges yet. To be consistent with the in-house model that is currently in use, we did an 80/20 random split of the data which left us with a training set of 5609 observations and a test set of 1434 observations. In order to maintain consistency between our results, we set a seed of 1905 before we split and ran our models.

Analysis

Classification Tree Model

The first model we built was a classification tree. We opted to split this tree using the Gini measure of impurity to decide optimal splits because it is one of the most popular splitting criteria. If Telco uses a different splitting criteria, we suggest these algorithms be rerun with the other splitting criteria in order to be consistent with Telco's in-house tree model. The graph of the classification tree is shown on page 2.

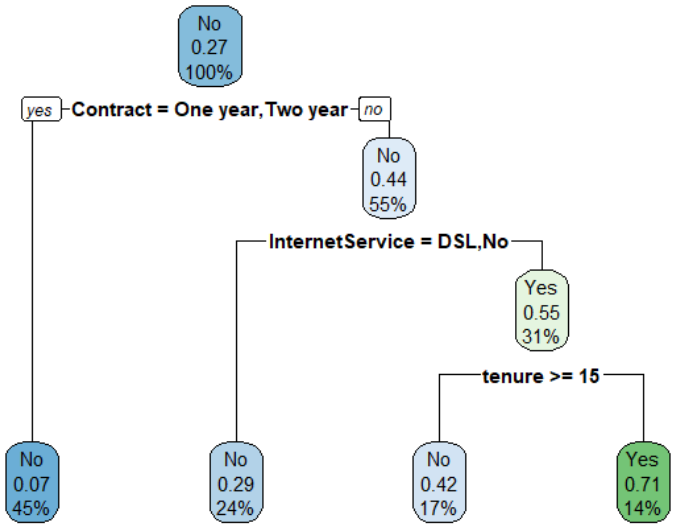

Figure 1: Classification Tree Model

In the classification tree model, the first split is based on whether the customer is signed to a service contract for at least a year or if the customer was on a month-to-month billing cycle. An interesting finding was that customers who are signed up for yearly contracts churn at a much smaller rate, approximately 7%, than either the overall churn rate or any other splits. The tree next split on whether the customer received fiber optic service or had DSL/no internet service. The last split in the tree was on whether or not a customer had been with Telco for 15+ years. This tree indicates that a customer with a month-to-month payment plan, fiber optic internet, and less than 15 years of tenure with the company has a 71% probability of churning.

Recursive Partitioning Tree Model

The second kind of tree our team built was a Recursive Partitioning Tree. The unpruned version of this tree was very complex, so we decided to apply an alpha of 0.0005 to pre-prune our results. The unpruned tree contains 22 inner nodes and 23 terminal nodes and can be viewed in the Appendix. The pre-pruned tree contains 7 inner nodes and 8 terminal nodes, which can be seen in Figure 2 on page 3. Similar to the classification tree, the first split is on Contract with month-to-month on one side and one and two-year contracts on the other. The next split is on Internet Service for both of Contract's child nodes. After that, the recursive tree deviates from the classification tree and increases complexity.

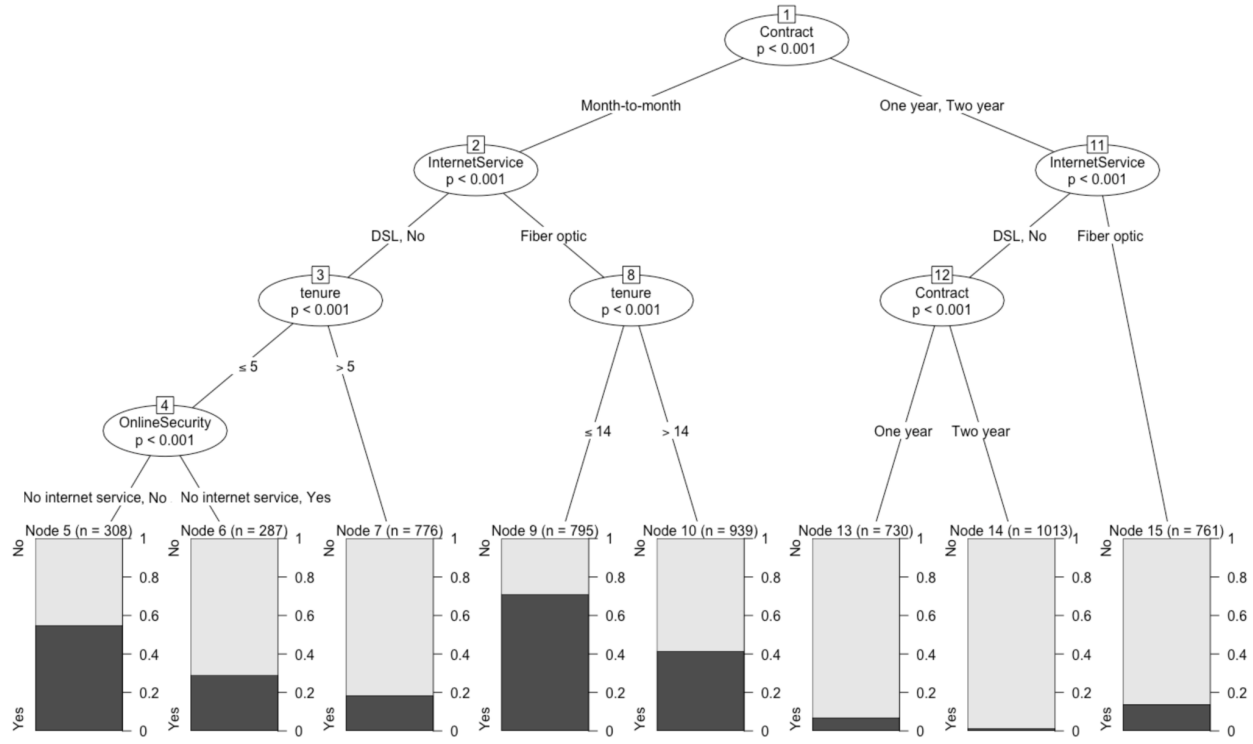

Figure 2: Recursive Partitioning Tree Model

Results

For both models, the misclassification rate and the ROC curve were calculated in order to compare the accuracy of their predictions. The table below presents the misclassification rates on the training and test data sets.

Table 1: Misclassification Rates On Training And Test Sets By Tree Model Built

Tree Model	Misclassification Rate on Training	Misclassification Rate on Test
Classification	21.09%	20.50%
Recursive Partitioning	20.56%	20.78%

As shown in Table 1, the classification tree had a misclassification rate of 20.50% on the test data set which was slightly better than the more complex recursive partitioning tree which had a misclassification rate of 20.78%. However, since Telco requested a binary decision (whether the customer churns or not), we calculated the area under the ROC curve to determine which tree produced a better combination of sensitivity and inverse specificity. In Figure 3 on page 4, we show the ROC curves for both models based on the predictions for the test data set.

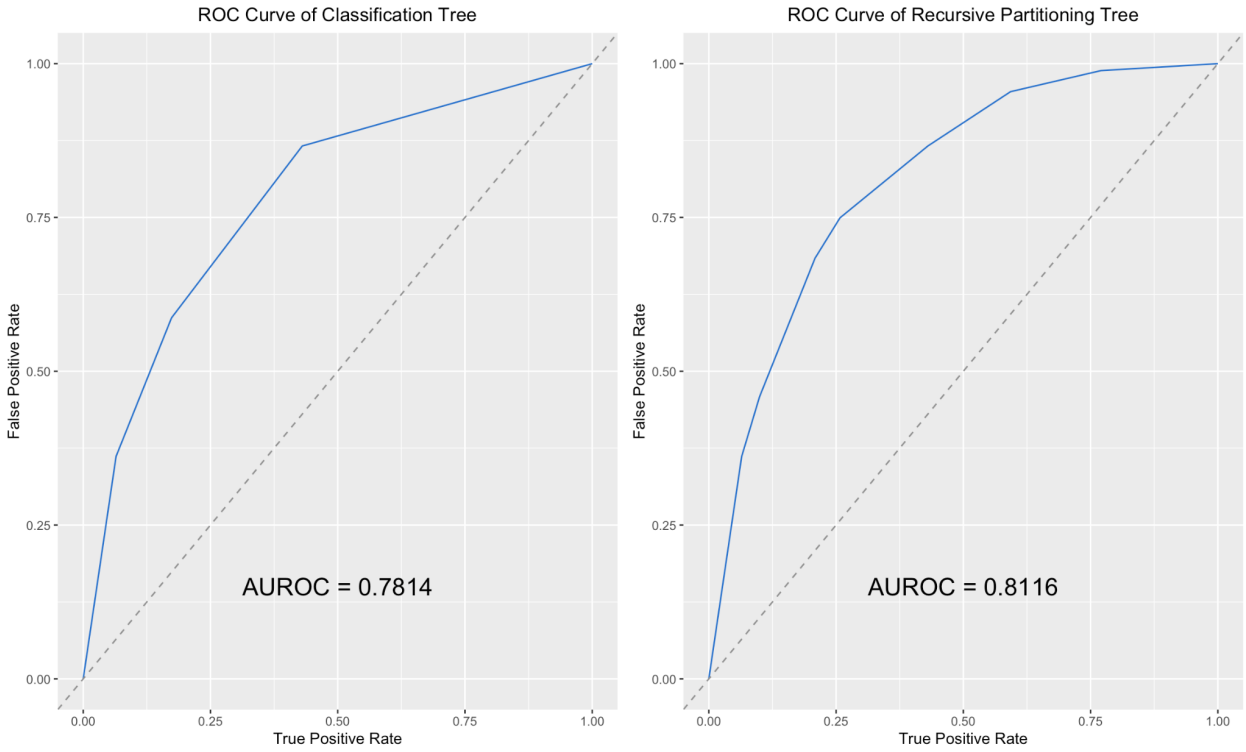

Figure 3: ROC Plots Of The Tree Models Built With Associated AUROC Values

The recursive partitioning tree's area under the ROC curve (AUROC) is 0.8116, which is better than the classification tree at only 0.7814. Based on all of these results, we recommend utilizing the classification tree because of its relative simplicity and lower misclassification rate on the test data. However, depending on the business needs and implementation method, the comparable misclassification rate and increased ROC curve does provide justification for utilizing the more complex recursive partitioning tree model.

Recommendations

We determined that the two most impactful variables on which to split a decision tree are the type of contract (month-to-month vs longer-term) and the type of Internet service (No and DSL vs Fiber Optic). As a result, going forward, we suggest that Telco:

- Prioritize longer-term (one- to two-year) contracts to reduce the likelihood of churning.
- Proactively market to existing customers with month-to-month contracts to encourage them to upgrade to long-term plans.
- Focus marketing efforts on fiber optic customers who account for 41.42% of monthly revenue in the training data. According to our model, these customers have a 55% probability of churning.

Conclusion

Our team built a classification tree and a recursive partitioning tree with the objective of predicting customer churn for Telco. The classification tree was built using the Gini measure of impurity and consisted of three optimal splits based on the various characteristics of a customer. We identified that the most impactful criteria for determining whether a customer will churn is

whether they are on a month-to-month payment plan or at least a yearly contract. The full recursive partitioning tree was much more complex, resulting in 7 inner nodes and 8 terminal nodes. We, therefore, opted to pre-prune the recursive tree resulting in seven inner nodes and eight terminal nodes. The misclassification rate on the classification tree was 20.50% on test data while it was 20.78% for the recursive partitioning tree.

After considering the misclassification rate on training and test data, as well as the AUROC, our team recommends utilizing the classification tree to predict customer churn due to its simplicity. However, our provided business recommendation is based on limited knowledge of Telco's current model and business needs. We suggest a follow-up analysis comparing Telco's current model and the models suggested in this report along with further investigation into Telco's business needs.

Appendix

```
[1] root
|  [2] Contract in Month-to-month
|  |  [3] InternetService in DSL, No
|  |  |  [4] tenure <= 5
|  |  |  |  [5] OnlineSecurity in No
|  |  |  |  |  [6] TechSupport in No: Yes (n = 251, err = 40.6%)
|  |  |  |  |  [7] TechSupport in Yes: No (n = 57, err = 35.1%)
|  |  |  |  [8] OnlineSecurity in No internet service, Yes: No (n = 287, err = 29.3%)
|  |  |  [9] tenure > 5
|  |  |  |  [10] PhoneService in No: No (n = 186, err = 28.0%)
|  |  |  |  [11] PhoneService in Yes
|  |  |  |  |  [12] tenure <= 22: No (n = 379, err = 19.5%)
|  |  |  |  |  [13] tenure > 22: No (n = 211, err = 7.6%)
|  |  [14] InternetService in Fiber optic
|  |  |  [15] tenure <= 14
|  |  |  |  [16] tenure <= 1: Yes (n = 187, err = 13.4%)
|  |  |  |  [17] tenure > 1
|  |  |  |  |  [18] MonthlyCharges <= 90.75
|  |  |  |  |  |  [19] OnlineSecurity in No
|  |  |  |  |  |  |  [20] MultipleLines in No: Yes (n = 226, err = 42.9%)
|  |  |  |  |  |  |  [21] MultipleLines in Yes: Yes (n = 171, err = 28.1%)
|  |  |  |  |  |  [22] OnlineSecurity in Yes: No (n = 49, err = 40.8%)
|  |  |  |  |  [23] MonthlyCharges > 90.75: Yes (n = 162, err = 19.1%)
|  |  |  [24] tenure > 14
|  |  |  |  [25] tenure <= 55
|  |  |  |  |  [26] SeniorCitizen in 0
|  |  |  |  |  |  [27] PaymentMethod in Bank transfer (automatic), Credit card (automatic), Mailed check: No (n = 261, err = 29.9%)
|  |  |  |  |  |  [28] PaymentMethod in Electronic check: No (n = 283, err = 48.4%)
|  |  |  |  |  [29] SeniorCitizen in 1: Yes (n = 266, err = 44.7%)
|  |  |  |  [30] tenure > 55: No (n = 129, err = 22.5%)
|  [31] Contract in One year, Two year
|  |  [32] InternetService in DSL, No
|  |  |  [33] Contract in One year
|  |  |  |  [34] OnlineSecurity in No: No (n = 192, err = 13.5%)
|  |  |  |  [35] OnlineSecurity in No internet service, Yes: No (n = 538, err = 4.5%)
|  |  |  [36] Contract in Two year
|  |  |  |  [37] SeniorCitizen in 0
|  |  |  |  |  [38] PaymentMethod in Bank transfer (automatic), Credit card (automatic), Mailed check: No (n = 894, err = 0.6%)
|  |  |  |  |  [39] PaymentMethod in Electronic check: No (n = 61, err = 6.6%)
|  |  |  |  [40] SeniorCitizen in 1: No (n = 58, err = 6.9%)
|  |  [41] InternetService in Fiber optic
|  |  |  [42] Contract in One year
|  |  |  |  [43] StreamingMovies in No: No (n = 111, err = 6.3%)
|  |  |  |  [44] StreamingMovies in Yes: No (n = 316, err = 23.1%)
|  |  |  [45] Contract in Two year: No (n = 334, err = 7.2%)
```

Figure 4: Recursive Partitioning Tree Graph

Homework Report Checklist

The team member(s) responsible for checking each item should enter their initials in the field next to each question. All items should be addressed before submitting the assignment with the initialed checklist attached.

Sections & Structure

Overview

JH	Is the overview concise?
JH	Does it provide context about the business problem? <Content>
JH	Does it briefly address your team's work, quantifiable results, and recommendations? <Action>
JH	Does it offer audience-centered reasons for recommendations? <Context>

Body Sections

JL	Does the report body include information on methods, analysis, quantifiable results, and recommendations?
JL	Is content grouped into appropriate sections (*methodology*, *analysis*, *results*, *recommendations*)?

Conclusion

JL	Does the report have a conclusion?
JL	Does the conclusion sum up the report and emphasize relevant takeaways?

Structure

JL	Does each major section have a heading?
JL	Are sections, subsections, and paragraphs organized logically for easy navigation?

Visuals

Introduction, Discussion, and Captions

JL	Is each visual introduced in the text before it appears?
JL	Is each visual close to where it is introduced?
JL	Does each visual include a title with the following information: type (*table* or *figure*), number, and a descriptive caption?
FM	Is each visual discussed and interpreted in the text?
JL	Are figures and tables numbered separately?
JL	Are table captions above the table? Are figure captions below the figure?

Visual Design

FM	Do figures/tables use audience-friendly labels rather than variable names?
FM	Are the visuals easy to interpret?
FM	Are the visuals appropriately sized?
JL	Do tables appear on one page (*not split between 2 pages*)?
JL	Are legends and axis labels included for figures?
JL	Are numbers in tables right aligned?
JL	Are the visuals designed well (*ex*: *re-created in Word or Excel*, *not blurry or stretched*,…)?

Document Design

Title Page Design

FM	Does it include a descriptive title?
FM	Does it state the team name, team members' names, and the submission date?

Table of Contents Design

FM	Does it list all the major sections of the report with corresponding page numbers?
FM	Do the page numbers and sections in the Table of Contents match the report?

Document Design for Entire Report

FM	Is a standard typeface (*Calibri*, *Arial*, *etc.*) used?
FM	Is the size of the body text between 10-12 pt.?
FM	Are headings and subheadings used to organize information?
FM	Are distinctive text styles (*bold*, *italic*, *etc.*) used to distinguish between heading levels?
FM	Are text styles for headings used consistently (*ex*: *all level-one headings are bold*)?
FM	Are all paragraphs an appropriate length (*fewer than 12 lines*)?
FM	Is white space used to indicate paragraph breaks?
FM	Are bullet lists used for a series of items and numbered lists to show a hierarchy?

Writing Style and Mechanics

Spelling and Capitalization

PO	Are spelling errors located and corrected?
PO	Is spelling consistent throughout (*no switching between acceptable spellings*)?
PO	Is capitalization used appropriately (*proper nouns*, *etc*.)?
PO	Is capitalization of words consistent throughout the report?

Grammar and Punctuation

PO	Are verb tenses used appropriately?
PO	Are marks of punctuation used appropriately?
PO	Is subject-verb agreement used in every sentence?
PO	Is the grammar checker updated and are underlined grammar issues addressed?

Writing Style

FM	Are all sentences in the report easy for your audience to understand quickly?
FM	Are most sentences written in active voice?
FM	Are idioms and vague words eliminated from the report?
FM	Are acronyms introduced before being used?
FM	Are well-written topic sentences included at the beginning of each paragraph?
FM	Are lists parallel?
FM	Is the appropriate point of view used when addressing your audience or describing team actions?