ⵕ **akaigraham** / **dsc-mod-3-final-project**

forked from learn-co-curriculum/dsc-project-template

Phase 1 Project Templates and Examples

⚖ View license

☆ **0** stars   ⵕ **167** forks

☆ Star	⊙ Watch ⌄

`<>` **Code**   ⑂ **Pull requests**   ⊙ **Actions**   ▥ **Projects**   📖 **Wiki**   ⛉ **Security**   ⃕ **Insights**

ⵕ master ⌄   ···

This branch is 30 commits ahead of learn-co-curriculum:master.   ⑂ Contribute ⌄   ⟳ Fetch upstream ⌄

🔴 **akaigraham** moved files to backup folder   ···   21 minutes ago   🕙 **69**

View code

☰ README.md   ✎

Phase 3 Final Project - Telecom Churn (Kai Graham)

README Outline

Within this README.md file you will find:

1. Introduction
2. Overview of Repository Contents
3. Project Objectives
4. Overview of the Process
5. Findings & Recommendations

6. Conclusion / Summary

Introduction

Build a classifier to identify whether a customer will "soon" churn and stop doing business with SyriaTel. Ultimate goal is to label at risk customers to enable the Company to "save" these customers via promotions or other outreach measures.

Repository Contents

1. `README.md`
2. `telecom_churn_classifier.ipynb` - clean jupyter notebook containing all code models
3. `telecom_churn.csv` - dataset used
4. `Untitled.ipynb` - working jupyter notebook containing rough, working, and in-process code

Project Objectives

Build a classifier to predict whether a customer will soon stop doing business with SyriaTel. Follow CRISP-DM Machine Learning process to explore dataset, prepare data for modeling, modeling, and post-model evaluation. We will also be focused on identifying which performance metrics will likely be best to evaluate our performance and ability to properly identify churning customers. Provide as an output a list of customers who are most likely to churn according to our best model to company.

Overview of the Process

Following CRISP-DM, the process outlined within `telecom_churn_classifier.ipynb` follows 6 key steps, including:

1. Business Understanding: Outlines facts and requirements of the project. Specifically, a classifier will be built and trained on various SyriaTel customer data to predict whether a customer will be labeled as a 1 (churn) or a 0 (non-churn customer). Understanding which customers are likely to churn, in addition to various patterns within the data should enable SyriaTel to perform more targeted customer outreach and hopefully relate various customer features with the strength of the customer relationship going forward.

2. Data Understanding: focused on unpacking all data that will be used in this classification problem (again primarily SyriaTel customer data). This section will focus on the distribution of our data, any imbalances within our target predictor, and the identification of which features are likely to impact or be associated with churn.

3. Data Preparation: Further preprocessing of our data to prepare for modeling. This includes splitting into training and test sets, encoding necessary columns, and handling any other data processing prior to modeling. This is also the section in which synthetic training data is created via SMOTE to help with class imbalance.

4. Modeling: this section trains and evaluates the performance of a number of machine learning models, primarily focused on decision trees, random forests, and XG Boosting algorithms

5. Evaluation: Final / optimal model is selected and final performance metrics of final model discussed and evaluated.

6. Deployment: Generate predictions on all data to provide SyriaTel with a list of customers that are at highest risk of churning based on our final / best classifier.

Findings & Recommendations

The best performing model we saw was our tuned XG Boosting algorithm, with an AUC of 0.865, f1 score of 78.6%, recall of 76% and overall acuracy of 94%. Looking at our final model feature importance, the most important features appear to be whether or not a customer is on an international plan, whether or not a customer is on a voice mail plan, and the number of customer service calls to date. Additionally, a list of 431 customers deemed "at-risk" of churning by our model. While these customers have already churned, model can be used going forward to generate a similar risk of existing customers with risk of churn. Recommended that Company begins targeted outreach / customer-saving metrics on this list of customers first.
dsfasfsfas

Conclusions & Summary

Through an iterative modeling and data preparation process, we were able to tune a model with ~76% recall, and overall accuracy of 94%. Throughout this process, recall and f1 score were favored over other metrics as the Company is likely not as concerned with false positives as customer-saving metrics targeted at this mis-labeled customers likely do not cost the Company much in comparison to potentially having that customer churn.

Releases

No releases published

Create a new release

Packages

No packages published

Publish your first package

Languages

● **Jupyter Notebook** 100.0%