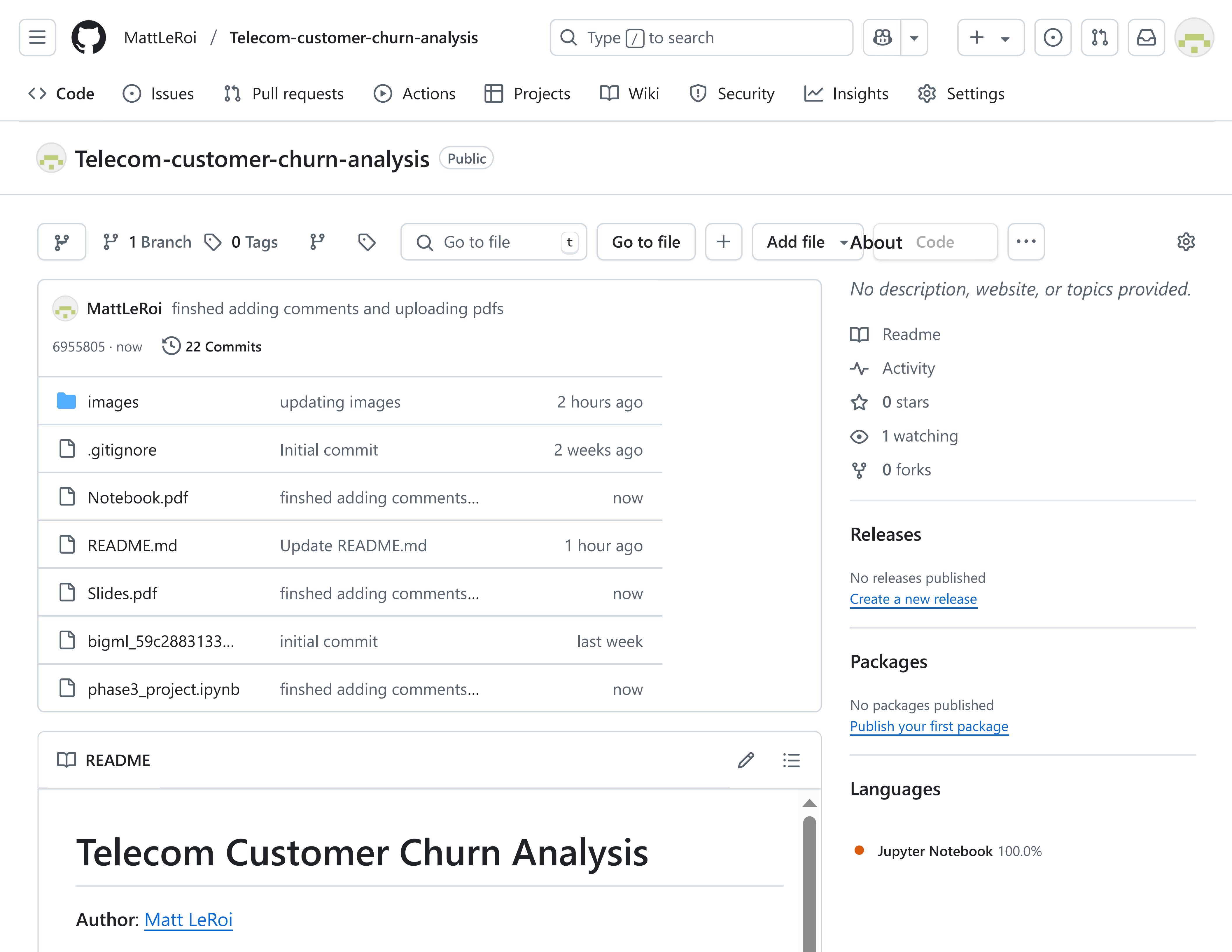


Telecom-customer-churn-analysis

MattLeRoi finshed adding comments and uploading pdfs

6955805 · now 22 Commits

images	updating images	2 hours ago
.gitignore	Initial commit	2 weeks ago
Notebook.pdf	finshed adding comments...	now
README.md	Update README.md	1 hour ago
Slides.pdf	finshed adding comments...	now
bigml_59c2883133...	initial commit	last week
phase3_project.ipynb	finshed adding comments...	now

About

No description, website, or topics provided.

Releases

No releases published

Create a new release

Packages

No packages published

Publish your first package

Languages

● Jupyter Notebook 100.0%

Telecom Customer Churn Analysis

Author: Matt LeRoi

Overview

In the telecom business, customers are generally on monthly or yearly contracts. The cost of acquisition of customers is high, so having sticky customers (customers that stick around once acquired and don't churn) is important. Therefore, predicting churn before it happens can reap huge financial benefits.

Business Understanding

The goal of this project is to create a model to predict churn in customers for SyriaTel, a telecom company. If customer churn can be identified before it happens, a retention strategy can be implemented before they churn. According to SyriaTel, the estimated lost profit due to churn is ~$240/customer and reaching out to a customer about to churn is ~⅓ effective, so the average profit from reaching out to a customer about to churn is ~$80. The cost of outreach is ~$20/customer. The cost of simply reaching out to all customers and retaining ⅓ of those about to churn would be a net loss of $28,000, so accurate prediction of customers likely to churn will be extremely valuable here. If one customer is correctly identified out of every four customers flagged for potential churn, this project will be net neutral. The model will seek a much higher bar than that, however, and will optimize for maximum profit.

Data

SyriaTel provided a subset of customer data. It included: 1 geographical location, 3 area codes, 3333 total customers, and 486 churning customers, with a 14.5% churn rate.

Modeling

Two models were created for this project: a logistic regression model and a decision tree classifier. These two were chosen because they can both handle making binary classification outcome predictions. Without knowing the type of relationship between the independent variables and the outcome (i.e. linear or more complex) beforehand, both models were created and compared. A baseline logistic regression model was created, then the data was scaled and oversampled (to combat the unbalanced data, since only 15% of customers churned). The baseline decision tree classifier was then created with various criteria and refined with hypertuning. During the tuning phase, a subset of the data was used for training and another set for validation. Once the parameters were set, the final evaluation was done using a final test set that wasn’t used during any of the training or tuning.

Evaluation

The goal of this project is financial, to increase profits for SyriaTel. The recall and precision scores are relevant, as we are trying to identify as many of the churning customers as possible (recall rate) while limiting the number of customers we reach out to unnecessarily (precision rate). To combine these two scores, I have taken the estimate provided by SyraiTel to directly calculate the actual financial impact to the company. Each correctly identified churning customer (True Positive, TP) is worth $80 on average, and every person SyriaTel reaches out to (True Positive plus False Positive, TP+FP) costs $20. The final evaluation criterion is then: $80TP + $20*(TP+FP), which is the total profit (or loss) of the experiment.

Conclusion

The final logistic regression model achieved a profit of $1300, correctly identifying 35 out of 101 churning customers, with 40 falsely identified churning customers. This is a positive result, literally, with a positive dollar value associated with it, but identifying roughly a third of the churning customers is not nearly as accurate as I would like.

The final hypertuned decision tree classifier model outperformed the logistic regression model, achieving a profit of $4180 after hypertuning, correctly identifying 83 out of 101 churning customers, with 40 falsely identified churning customers.

Limitations

The available data covers a small number of customers in one geographical area and only includes basic usage and account data. It appears that there is enough to be useful, but more detailed data could be extremely helpful to refine the model further. Also, the financial evaluation is based on rough estimates, so further experimentation can refine the evaluation criteria.

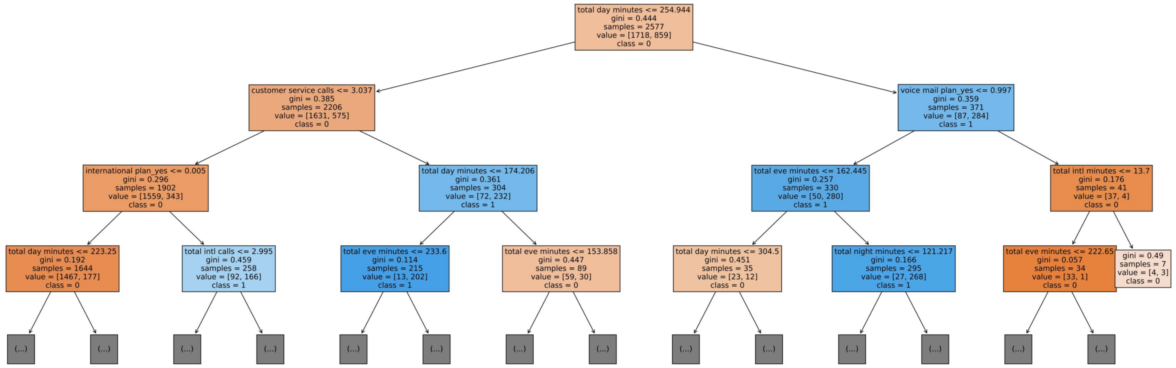
Recommendations

The model provided should yield positive results. Broader data, however, covering more of SyriaTel’s customers and containing more varied and detailed information, should help refine the model. Also, a few insights from how the classification tree was built:

Customers with high usage and no voice mail plan churned at a very high rate, ~90%. I recommend looking into this group further. Is this line only used for a specific purpose and then closed? These could be high value customers since their usage is high, so the rewards for solving their high churn rate could be great.

Customers with low usage and high customer service calls churned at a high rate, unsurprisingly. Looking for patterns in customer service calls may lead to better customer engagement with more usage and greater retention.

Finally, customers with low usage, few customer service calls, and no international plan stayed at a relatively high rate. Generally speaking, customers with higher usage churned at a higher rate in several points in the tree. It may be worth looking into pricing strategies that reward greater use rather than a flat per-minute rate. This could help retain the high value customers and prompt other customers to use the service more.



For More Information

See the full analysis in the Jupyter Notebook and presentation.

For additional info, contact Matt LeRoi at mcleroi@gmail.com

```
├── images
├── README.md
├── Slides.pdf
```