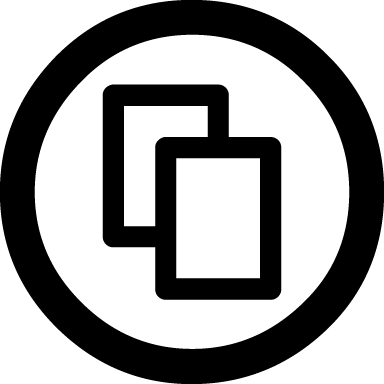
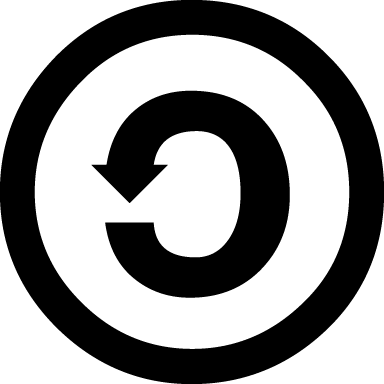
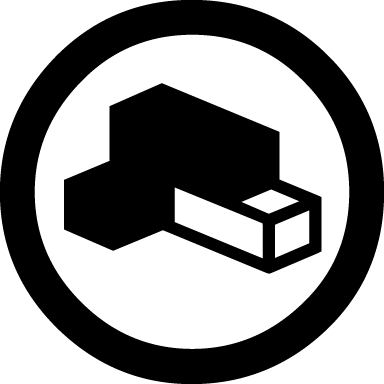
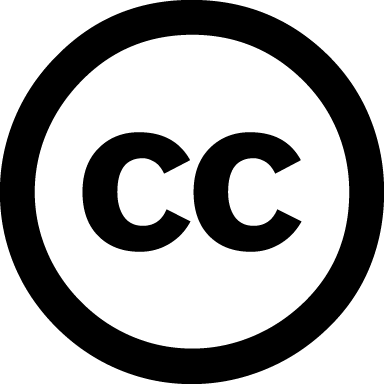
# The Machine Learning Canvas (v0.4) Designed for: Telecom company churn prediction Designed by: Xinyan Lyu Date: May. 2019 Iteration: 1 /month .

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Decisions**  How are predictions used to make decisions that provide the proposed value to the end-user?  Every Month:  Compute predictions for all customers with available info.  Prioritize and retarget on customers with high predicted possibility of churn, but haven’t really churn now.  Review monthly and identify at-risk customers, and look deep into their reasons of churn. | **ML task**  Input, output to predict, type of problem.  «Whether a customer would churn? »  Input: customers’ tele package usage data, demographic data, payment data, and customer service data.  Output: risk of churn( 0 or 1, 1 represent churn)  -> classification task | **Value Propositions**  What are we trying to do for the end-user(s) of the predictive system? What objectives are we serving?  Make more accurate predictions of the churn of specific customers. Measure the risk level of the whether a customer would churn in the future. Understand the correlation between payment, engagement data(phone call and message data, etc.) and the risk of churn. | **Data Sources**  Which raw data sources can we use (internal and external)?  Customer Churn Prediction dataset from Kaggle.  (https://www.kaggle.com/c/customer-churn-prediction/data) | **Collecting Data**  How do we get new data to learn from (inputs and outputs)?  Every month request data on:  Telecom package plan usage, payment, churn data from Telecom internally . |
| **Making Predictions**  When do we make predictions on new inputs? How long do we have to featurize a new input and make a prediction?  Every month we make predictions for all customers(renewing all information available once a month, because, generally speaking, most people pay telecom company once/month and their mobile package plans are monthly, or telecom company calculate payment for a customer once/month). | **Offline Evaluation**  Methods and metrics to evaluate the system before deployment.  The train and test datasets have already been split.(train: test is about 1:2)  ROC and AUC  Focus on FN and FP, especially on FN, which means not successfully predicted churn. Based on this evaluation, keep tuning the model, but also mind overfitting problem. |  | **Features**  Input representations extracted from raw data sources.  customers’ tele package usage data, demographic data, payment data, and customer service data.  Such as :  number vmail messages, total day charge, total night call minutes, number of customer service calls, etc. | **Building Models**  When do we create/update models with new training data? How long do we have to featurize training inputs and create a model?  Should keep long-term data, which is helpful to track the history data , and history churn rate.  Update model every month(with new data available). |
|  | **Live Evaluation and Monitoring**  Methods and metrics to evaluate the system after deployment, and to quantify value creation. | Percentage of customers who churn in the next month or next few months.  Track longer term customer data in the past, because customer churn may result from problem associated with payment and service in the past few month. |  |  |

[**machinelearningcanvas.com**](http://www.machinelearningcanvas.com) **by Louis Dorard, Ph.D.** Licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.****