

Predict Client's Churn

Sparkify is a digital music service similar to **Netease_Cloud_Music** or **QQ Music**. Many of the users stream their favorite songs in Sparkify service everyday, either using free tier that places advertisements in between the songs, or using the premium subscription model where they stream music as free, but pay a monthly flat rate. User can upgrade, downgrade or cancel their service at anytime.

This is a Customer Churn Prediction Problem , there are so many similar projects, such as WSDM - KKBox's Churn Prediction Challenge competition from Kaggle, and a few helpful links are follows:

- Customer Churn Prediction using Machine Learning (How To)
- Prediction of Customer Churn with Machine Learning
- Customer Churn Prediction and Prevention
- Hands-on: Predict Customer Churn

So, our job is data mining the customers' data and implement appropriate model to predict customer churn as follow steps:

- Clean data: fill the nan values , correct the data types, drop the outliers.
- EDA: exploratory data to look features' distributions and correlation with key label (churn).
- Feature engineering: extract and found customer-features and customer-behavior-features; Implement standscaler on numerical features.
- Train and measure models: I choose logistic regression, linear svm classifier, decision tree and random forest classifier to train a baseline model and tuning a better model from best of them. It is worth mentioning that this data is unbalanced because of less churn customers, so we choose f1 score as a metrics to measure models' performance.