Churn Analysis

Ngan Bui

Topic

- Predicting Churn within paid customers at Zoosk
- Identify customers who cancel their subscriptions within 30 days
- The data includes customers with their interactions on the site within 14 days of subscribing
  - Can we use user interactions on the site to predict whether or not they will cancel?

Acquiring the Data

- Demographic data were easy to pull
- For interactions, stored in Hadoop
  - Only have logs where each interaction between 2 users is a record
  - So aggregating interaction at a per user level, 14 days after subscription took a lot of SQL
  - However, if I pull the data whole dataset for the last year, I could end up with millions of rows
    - Decided to take a sample (10,000 users) as a trial run

The data - Interaction on the sites

- Cancel 30 - whether or not the user cancels within 30 days
- Outgoing & incoming messages within 14 days
- Outgoing & incoming profile views within 14 days
- SmartPick matches within 14 days
- Carousel matches within 14 days
- Message sentiment sum for each user at the end of 14 days
- Total logins within 14 days
- How many days it takes for them to subscribe
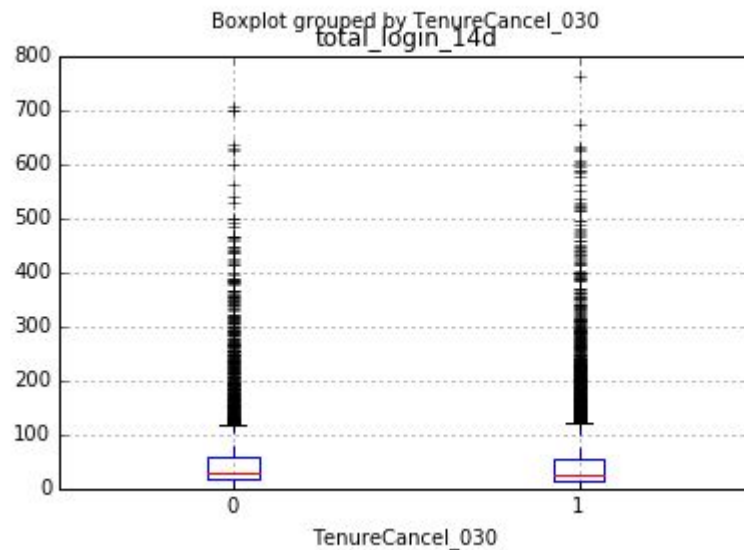- Whether or not the user had subscribed before

Demographics Data

- Gender
- Age
- Ethnicity
- Education
- Height
- Platform
- Marketing Channel
- Month of subscription
- Payment method

 -> 70 features (including dummy variables)

Total logins vs. Churn

Classification Model

- KNN
  - N_neightbors = 5 yield the best results
- Logistic Regression
- Naive Bayes
- Random Forest

Results: Beats the null accuracy rate by 0.1 percentage point

Best model -> Select K Best (k=4), then Random Forest

N_estimator =300, max_features = 4

Conclusions

- In the future, the project could be extended to include a much larger dataset
- While the model barely beats the null accuracy rate, it means that it is possible to predict churn
- Future solutions:
  - Dig through the data and add more features to see if there are any meaningful features
  - Extend 14 days after subscription to 30 days
  - Add notifications (email, push) data