Successful
Enterprise Social Collaboration

Using Data Science to Solve the Problem of Generating Successful Adoption

Micah Markman

DAT8 - Fall 2015 - Final Report

Successful
Enterprise Social Collaboration

Using Data Science to Solve the Problem of Generating Successful Adoption

The Challenge

Enterprise Social Collaboration blew up as technology that could revolutionize the knowledge workplace in the last 3 years. A host of pure-play start-ups brought the simplicity and usability of Facebook and LinkedIn into the workplace in an effort to revitalize interest collaboration software and deliver the value promised by the intranets and corporate knowledge portals that originally came into being in the mid-to-late 2000's. These companies promised that replicating the experience of Facebook in the enterprise would generate the user adoption that is critical to actually deriving the return on investment these solutions can deliver. Despite these promises, many implementations have suffered the same adoption challenges seen in the prior generation of intranet projects. My theory is that these implementation are still failing to execute on key organizational change management best-practices; specifically failing to find ways to bring users into the system and demonstrate the value in a way that causes minimal pain to each user and may even be fun.

How can data science help solve this problem of getting users to adopt these social collaboration platforms and support the overall user behavior change process? Up until now, all of the efforts that I've seen to support and drive changing user behavior has been done based on anecdotal evidence and historical best practices. For example, we need to do lots of user training sessions because that's what is done. Another example that has taken off is "gamifying" usage of these systems by rewarding users with points for completing a variety of usage activities within the system. The theory is that if you can get a user to develop certain habits when they first start using the system, the will continue even after the user tires of "game". However, the rules that dictate how users earn points have been generated based on anecdotal evidence and theories about what will work. My belief is that we can use

data science to understand the actual behaviors and attributes of users who become successful drivers of value and the create materials and functionality in the platform to drive all users follow the path laid out by early successes. Specifically, I believe that if we can train a model to predict whether a user will "adopt" the platform, then we look at what the model believes impacts this classification and use this data to make more targeted interventions with users as well as better "games" to build formative habits to get more users to follow the path of the successful adoption.

Focusing the Problem and Hypothesis

Source Data

The core of the data available for this lives in a couple of tables in a couple of SQL databases:

User

Column	Data Type	Notes / Description
userid	bigint	unique id for user
creationdate	bigint	easily convertible to timestamp for readability
userenabled	bool	0 no longer enabled 1 enabled
usertype	bool	real user (0) vs. placeholder for invited user (1)

UserProfile

Column	Data Type	Notes / Description
userid	bigint	reference to user.userid
fieldid	bigint	which profile field this row is for (reference to profilefield
levelid	bigint	level of sharing of profile data element (reference to profileseclvl
value	varchar	actual user profile field value

Pre-processing and Scrubbing

1. Limit to users have joined system between January and April 2015; sufficient count and this way any other variables outside data in system can be controlled for (calendar cycles, other activity within organization, etc).
2. Normalizing data to show the following per user (big effort):
   2.1. UserId
   2.2. Creation Date
   2.3. Location
   2.4. Organization
   2.5. Per Month Data
   2.6. # Logins
   2.7. # "Small" Activities
   2.8. # "Large" Activities
   2.9. # Group Memberships
   2.10. # Social Relationships
   2.11. Most Active Group # "Small" Activities
   2.12. Most Active Group # "Large" Activities

Data Exploration

1. ~48% of users drop to 0 activity after 6 months
   1.1. This includes users who leave the organization completely so this "end-state" is over-reported; trying to find a way to clean these users out of data.
2. ~21% of users are "Contributors" after 6 months

Intuitions based on playing with data
3. Thus far, no clear pattern in first month personal activity; some users do a lot and then go quiet; others start with little activity and then go nuts. This lead to re-definition of activity into "Small/Large"; haven't had a chance to really understand this.
4. Membership in an "Active" group that adopts tool seems to correlate with continued participation; do we need to further engineer this feature? A

significant percentage of "Contributors" are members of top 50 groups by activity.

5. Anecdotal evidence / theory that number of relationships built with active users leads to activity (looking to engineer a feature to represent this).

Feature Selection and Data Transformation

Modeling Process

I have realistically done very little modeling thus far. Here's my plan:
Logistic Regression (started on this and have some preliminary results; nothing as good as null model yet).
Decision Trees (this is the model that I have most hope for)
Ensembling

Challenges and Successes

1. Scrubbing training attendance data has proven to be more work than currently have time for; have to eliminate this feature but want to bring back.
2. Realized that there were a lot of users that joined but didn't really ever actively engage and fell off very quickly. These seem to introduce a lot of noise/bias into the system. May want to rerun some models after excluding these users.

Conclusions

Business Applicability and Next Steps

There are some immediate takeaways of adjusting our current gamification rules as well as general outreach to users