Predicting User Product Adoption Project Summary

Take-home Challenge

Nicole Quevillon

February 2025

Findings:

Our best performing model was a Gradient Boosting Classifier with a 70% precision rate and 78% accuracy.

Interestingly, the most important feature by far was how long it had been since the user created their account. This raises some important questions about the users included in the data. Does this data include all users who have signed up, or only those still active? This may be biasing the results if this data excludes users who signed up but have since deactivated their accounts.

	Importance
time_since_creation	0.778551
opted_in_to_mailing_list_True	0.083307
enabled_for_marketing_drip_1	0.060272
creation_source_PERSONAL_PROJECTS	0.019216
invited_True	0.017918
creation_source_SIGNUP_GOOGLE_AUTH	0.014076
creation_source_SIGNUP	0.013402
creation_source_ORG_INVITE	0.013258

If this does include all users who have signed up during the study time frame, this brings up some interesting questions. What has changed over this period that could have such a drastic impact on user adoption? Has the onboarding flow been updated? Has marketing's user targeting changed? Have there been significant changes to UX/UI?

These finding suggest that the likelihood of user adoption has decreased drastically over time and, if not caused by data bias, definitely needs to be explored in more detail.

Possible Project Improvements:

There was a tight turn around on this project, so there are some areas that could be improved:

- Users were flagged as 'adopted' if they visited at least 3 times in any week, defined as set Sunday to Saturday periods. This logic could be improved by looking for any consecutive 7 days in where they visited at least 3 times using a rolling window function.

[GitHub Link]