Predicting User Adoption: A Concise Analysis

This report details an analysis identifying factors that predict future user adoption, defined as logging into the product on three separate days within at least one seven-day period.

Approach

My methodology involved:

1. **Data Preparation**: Loading user metadata and login history, converting timestamps to datetime objects, and sorting data.
2. **Defining "Adopted Users"**: Users were flagged as "adopted" if they had at least three unique login days within any seven-day rolling window. This status was merged into the main dataset.
3. **Feature Engineering**: New features like creation_year, creation_month, creation_day_of_week, and a has_invited binary flag were derived. creation_source was one-hot encoded, and missing values were handled.
4. **Modeling**: A Random Forest Classifier was used for prediction and to determine feature importances. The data was split for training and testing, and the model's performance was evaluated.

Key Findings and Predictive Factors

The Random Forest model highlighted several strong predictors of user adoption:

- **Organizational Affiliation (org_id)**: By far the most significant predictor (over 55% feature importance), suggesting that a user's organization is critical. This could be due to varying organizational structures, internal promotion, or specific team use cases.
- **Temporal Factors (Creation Month and Day of Week)**: Highly influential (approx. 32% combined importance), indicating that seasonal trends or weekly patterns in sign-ups correlate strongly with future adoption. Users signing up during specific periods might be more engaged.
- **Creation Source**: Users joining via an ORG_INVITE (invited as a full member to an organization) showed a higher adoption rate compared to other sources, though this factor was less dominant than org_id and temporal aspects.

Factors Considered / Did Not Pan Out

- **Marketing Opt-ins**: Features like opted_in_to_mailing_list and enabled_for_marketing_drip showed low predictive power (each below 3% importance), implying these general marketing flags are not strong direct indicators of adoption.
- **Individual Invitation**: Whether a user was invited by another individual user (has_invited) also proved to be a weak predictor, distinguishing it from the stronger effect of *organizational* invitations.

Further Research and Data

To deepen understanding and refine predictions, I recommend:

- **Deep Dive into org_id**: Investigate characteristics of high-adoption organizations (e.g., industry, size, specific product use cases).

- **Detailed User Activity Logs**: Analyze specific feature usage, interaction frequency, and session durations to understand the pathway to adoption.
- **A/B Testing Onboarding**: Experiment with different onboarding flows based on creation_source to optimize initial user experience.
- **Additional User Data**: Incorporate demographic information or user roles if available.
- **Customer Support Data**: Explore the impact of support interactions on retention and adoption.

By prioritizing organizational context and understanding user acquisition timing, the product team can develop more targeted strategies to drive user adoption.