USER CHURN PROJECT | ML Model Results

Prepared for the Waze Leadership Team

Issue / Problem

Waze's data team is building a predictive model to identify users at risk of monthly churn (uninstalling or ceasing app use). This milestone delivered the first production-ready classifier, tested on held-out data and paired with a business-aligned decision threshold so retention teams can act with confidence.

This report summarizes Milestone 6 and its impact for any future development.

Response

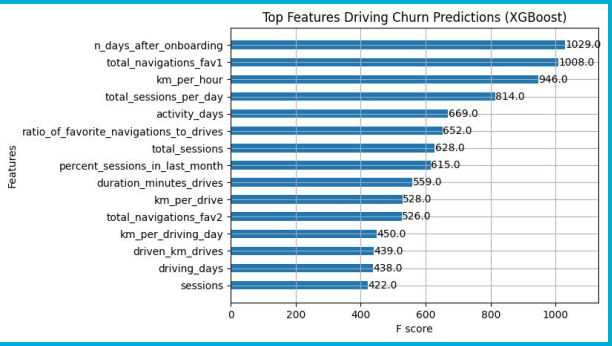
- We trained and compared two ensemble classifiers—**Random Forest** and **XGBoost**—using a three-way split (train/validation/test). A separate validation set enabled objective model selection; only the final champion was evaluated once on the untouched test set to estimate future performance.
- We optimized first for recall (to minimize missed churners) while tracking precision, ROC AUC, and PR AUC.
- We then selected an operational decision threshold from the Precision–Recall curve to align with Waze's outreach strategy.

Key Insights

- **Champion model**: XGBoost outperformed Random Forest on recall.
    - **Validation**: recall ≈ 65%, precision ≈ 34%, ROC AUC ≈ 0.75.
    - **Test**: recall ≈ 61%, precision ≈ 31%, ROC AUC ≈ 0.72, PR AUC ≈ 0.35.
    - Versus the dummy baseline (recall ≈ 16%), the model captures roughly 3–4× more churners.
- **Operational threshold**: Using the PR curve, we adopted a balanced cutoff near 0.58 to meet a recall floor of ~50% with ~35% precision and higher accuracy than more aggressive settings. If outreach is low-cost (email/banners), an optional light-touch tier at 0.40 raises recall to ~73% (precision ~28%).
- **Behavioral drivers**: Early engagement and consistent use are critical. Top signals include days since onboarding, recent session intensity, favorites usage, and engineered ratios such as percent of sessions in the last month and favorite navigations per drive.

Impact

- **Retention leverage:** With recall between ~50–60% (depending on threshold), the model materially reduces missed churners and provides a ranking to prioritize outreach.
- **Cost control**: Threshold selection lets Marketing trade precision for coverage. A two-tier approach (≥0.58 high-touch; 0.40–0.58 light-touch) balances ROI and user experience.
- **Transparency**: Confusion matrices quantify false positives/negatives; feature importance clarifies why users are flagged, informing onboarding and re-engagement strategies.



Recommendations

- ➔ **Deploy the champion and threshold**: Persist the fitted XGBoost model and the ~0.58 operating cutoff; score users on a daily/weekly cadence.
- ➔ **Tiered outreach**: High risk (≥0.58) receives prioritized, higher-touch actions; medium risk (0.40–0.58) receives lighter-touch nudges.
- ➔ **Onboarding focus**: Increase engagement in the first 60–90 days (tutorials, prompts, etc).
- ➔ **Measure and tune:** Track precision, recall, overall flag rate, and campaign ROI; re-estimate the threshold and retrain on a regular basis
- ➔ **Data enrichment (next iteration)**: Explore additional in-app interaction signals to raise precision without sacrificing recall.