TEAM 18 / DATA SCIENCE FINAL PROJECT

Beyond Negative Reviews on Steam



Which dissatisfied players should game developers prioritize?

Predict engagement. Use complaint themes to guide responses.

Data and prediction target

16.4M

negative reviews

From 113.9M Steam reviews
covering 8,183 games

Kaggle reviews + SteamDB metadata
One record = one player–game review

INPUTS AVAILABLE AT POSTING

Prior playtime, reviewer characteristics, review timing and NLP complaint features

ENGAGEMENT TARGET

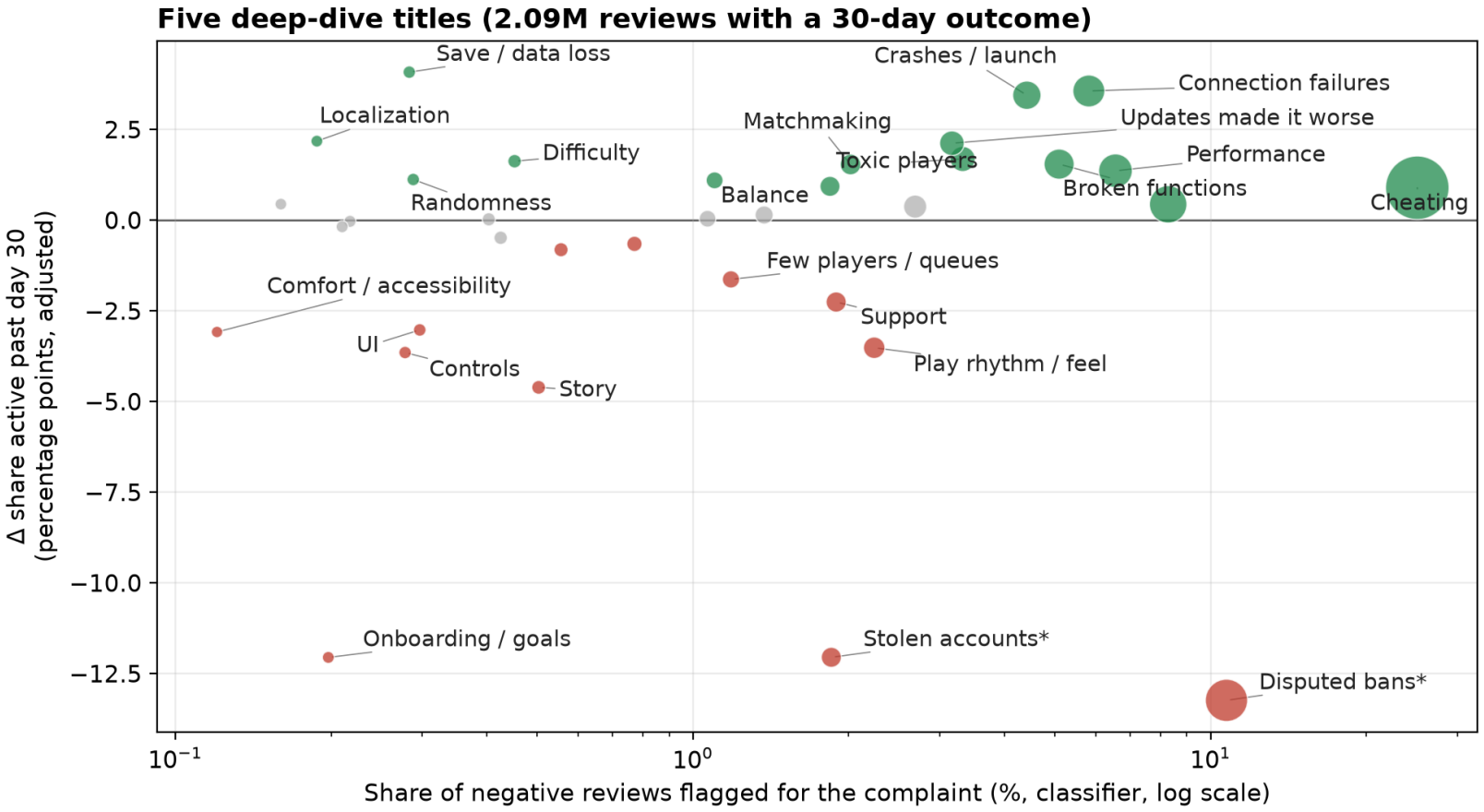
D30 engagement proxy from historical data
Unequal follow-up limits interpretation

TRAIN	VALIDATE + CALIBRATE	TEST
2010–2021	2022	2023

Historical snapshot: the target does not confirm activity on exactly day 30.

Engagement varies by complaint type

35 of 78 complaint factors met the threshold for use at scale



–12 pp

Onboarding complaints
Lower engagement in five deep-dive games

+1 to +7 pp

Technical complaints
Higher engagement

Red = lower share active past day 30, green = higher (90% CI excludes 0); grey = not significant. Bubble area ∝ number of flagged reviews.
Labels: |effect| ≥ 1 pp or flagged share ≥ 10%.
* Bans may partly limit play mechanically; the stolen-account association is not confirmed outside the deep-dive titles.

Adjusted associations, not causal effects

LightGBM + NLP performs best

10-fold model comparison, 2010–2021

Model / features	Average precision
Logistic / structured	0.8012
L2 logistic / + NLP	0.8194
LightGBM / structured	0.8226
LightGBM / + NLP	**0.8457**

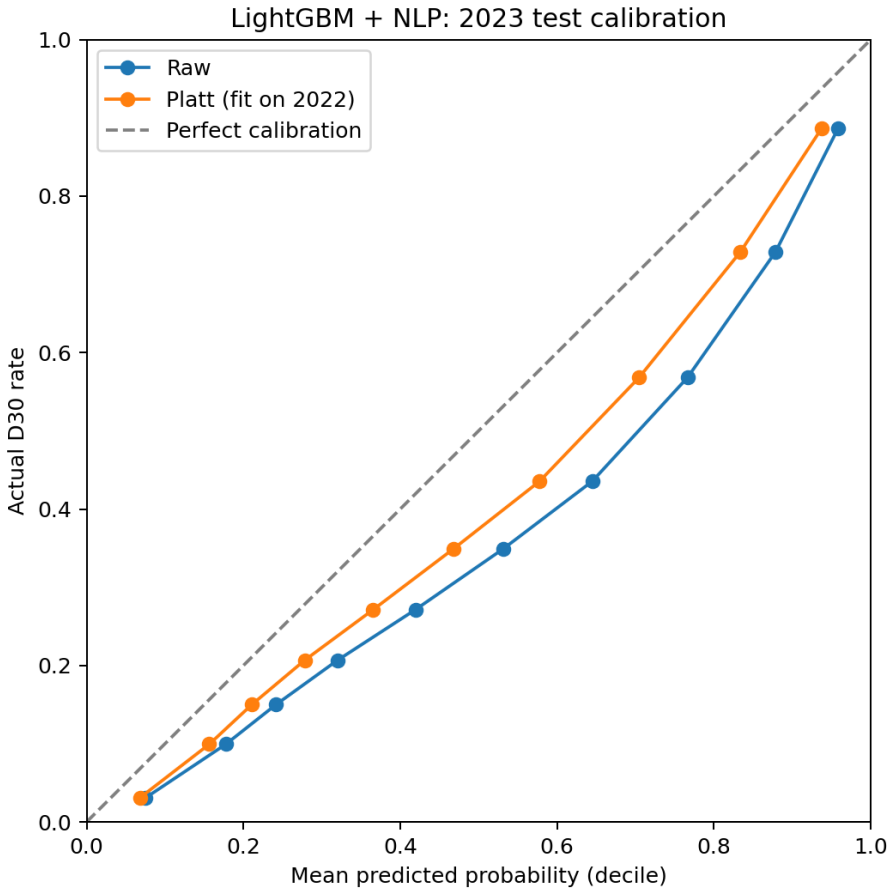
2023 TEST: 2.15M REVIEWS

0.824

ROC-AUC

88.7%

Engaged in top-scored 10%
2.38× lift over test baseline



Mean prediction 46.0% vs. observed 37.3%

Probabilities still overestimate by 8.7 pp

Four segments guide review prioritization

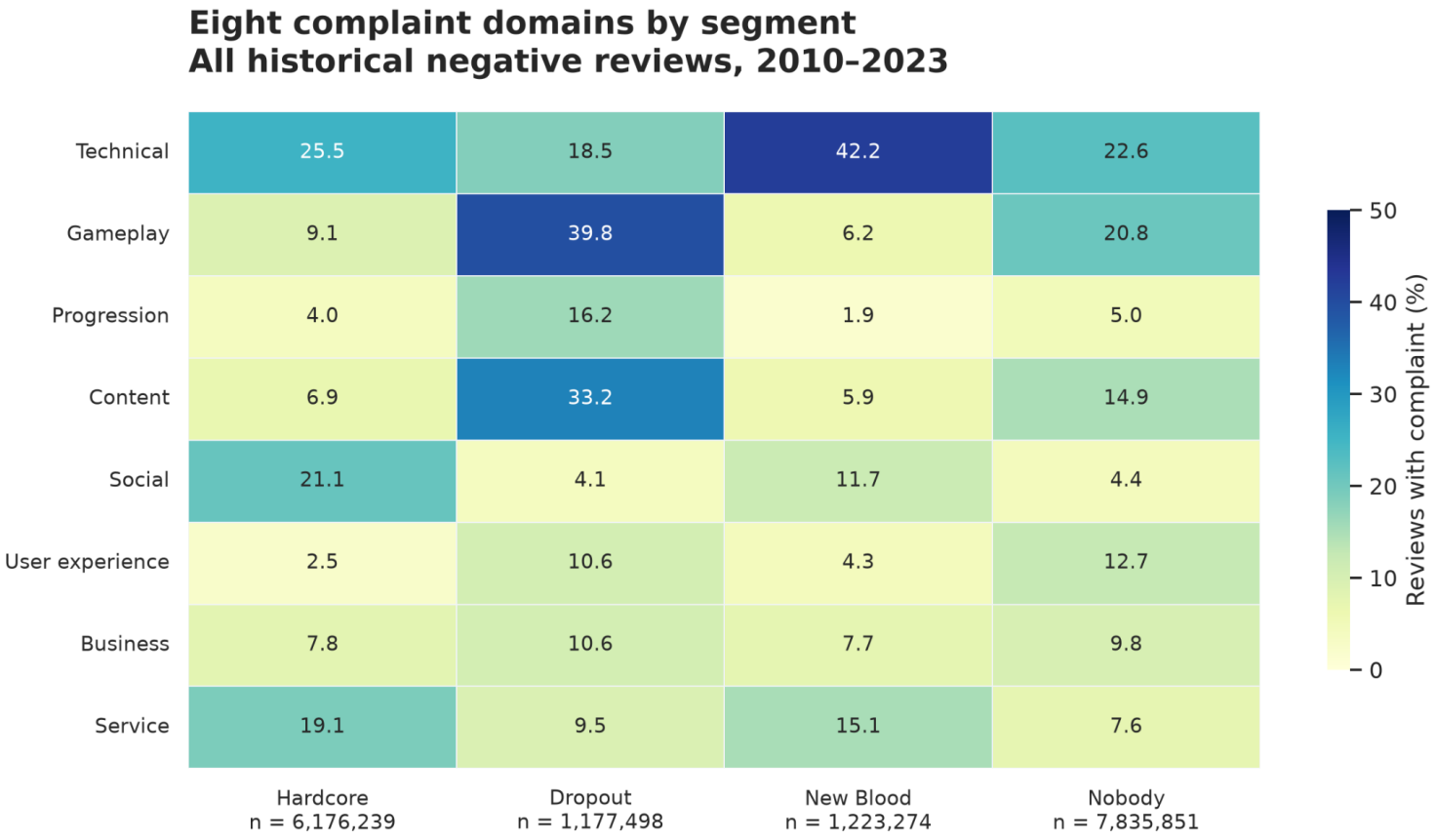
2023 test set, 2,154,782 review-level records

Segment	Prior playtime	Predicted engagement	Share
Hardcore	High	High	**36.06%**
Dropout	High	Low	**6.59%**
New Blood	Low	High	**8.94%**
Nobody	Low	Low	**48.41%**

High playtime: > 11.78 hours
High predicted engagement: at or above each period’s 55th percentile

Relative profiles, not confirmed churn. The “Nobody” label does not imply low-value feedback.

Targeted responses need a measured pilot



Each cell uses all records in its segment as the denominator. Domain = at least one factor flagged. Complaints can overlap.

Dropout

Investigate gameplay and progression dissatisfaction

New Blood

Resolve technical friction in the early experience

Hardcore

Use service and social feedback to improve ongoing play

Pilot targeted responses against current practice

Measure engagement and support costs. Recalibrate scores and keep human review.