

SPEC2MODEL Challenge — Submission Sheet

Section 1 — Team Information

Team Name	Team Hiatus
Problem Chosen	B
Member 1	Devi Prasad Tripathy
Member 2	Roshan Choudhary
Member 3	Swayam Agarwal
Contact Email	tdevi.prasad1@gmail.com

Section 2 — Data Strategy

Used Provided Dataset Only. No external data sources were used.

Section 3 — Data Cleaning & Preprocessing

- 1 Missing values filled using median imputation.
- 2 Reduced impact of noisy features and dropped irrelevant columns.
- 3 Handled class imbalance using stratified methods and class weights.
- 4 Converted categorical features into numeric encoding (0/1/2).

Section 4 — Feature Engineering

Rank	Feature	Why It Matters
1	feature_used_pct	Indicates engagement level
2	subscription_duration_days	Shows long-term behavior
3	days_since_last_login	Strong inactivity signal

New Features: price_pain_signal, app_flow_rating, crashes, alt_signal, disengagement_score.

Dropped ~15 least important features.

Observed mild correlations between ticket-related columns.

Section 5 — Model Selection & Training

Model Tried	Result	Reason
XGBoost	Lower	Less stable on categorical data
LightGBM	Lower	Early stopping issues
CatBoost	~57%	Best performance

Final Model: CatBoost

Hyperparameters: K-Fold=5, Early Stopping, Class Weight≈50

Validation Strategy: Stratified K-Fold Cross Validation

Best Validation Score: ~57%

Section 6 — Honest Reflection

- 1 Early stopping for LGBM & XGB did not work well.

- 2 Could improve K-fold tuning and feature evaluation.
- 3 Hardest part was understanding correlations and limited dataset familiarity.
- 4 With more time: refine early stopping and engineered features.

Section 7 — Team Collaboration

Member	Contribution
Devi Prasad Tripathy	Feature Engineering, Parameter Tuning
Roshan Choudhary	Data Cleaning & Preprocessing
Swayam Agarwal	Encoding, Prediction, Report Analysis

Section 8 — External Data Declaration

Yes — No external data used beyond provided dataset.