

PayPulse

Propensity-to-Pay Prediction for Healthcare Accounts

PayPulse is a machine learning system that predicts which healthcare accounts are likely to pay. It uses an ensemble of 4 models trained on historical payment data, with smart frequency encoding to handle high-cardinality features like Zipcode and Customer without memory issues.

Ensemble Models

XGBoost

LightGBM

RandomForest

CatBoost

How It Works

DAILY FLOW

SQL prepares accounts → Python scores them → SQL moves predictions to production → Python captures outcomes from old predictions to grow the training dataset.

SELF-LEARNING

The feedback loop tracks predictions for 30–120 days, captures actual outcomes (paid / not paid), and adds them to training data. Weekly retraining on this growing dataset makes the model smarter over time.

SMART ENCODING

Frequency encoding handles high-cardinality features like Zipcode and Customer efficiently, avoiding the memory explosion that one-hot encoding would cause with thousands of unique values.

Daily Pipeline

1

SQL Prep

Prepare eligible accounts

2

Python Score

Ensemble model scoring

3

SQL Deploy

Move to production table

4

Feedback

Capture outcomes for retraining

Output

PAYPULSE_SCORE TABLE

Each account receives a probability score (0–100%) and a prediction label stored in the PayPulse_Score table for business use.

85%

Can Pay

32%

Might Not Pay