Executive Summary: Energy Customer Churn & Price Sensitivity

Random Forest churn model with pricing-feature engineering and threshold tuning

Updated: 14 Jan 2026

Objective & Data

- Predict customer churn and assess whether pricing behaviour is a key driver.
- Datasets: client data + price history (time series by customer id).
- Churn is imbalanced: 9.72% churners (1,419) vs 90.28% non-churners (13,187).

Approach

- EDA: validated data types, missingness hotspots (channel_sales ~25.5%), and heavy-tailed usage/margin distributions.
- Feature engineering: lifecycle, consumption ratios, forecast error; aggregated pricing (latest/mean/volatility, spreads, change over time).
- Model: Random Forest (class_weight='balanced') with one-hot encoding and median/mode imputation; evaluated using ROC-AUC, precision, recall, F1.

Model Results

Scenario	Accuracy	Precision	Recall	F1	ROC-AUC
Baseline (all non-churn)	0.9028	0.0000	0.0000	0.0000	—
Random Forest @ 0.50	0.9100	0.8889	0.0845	0.1543	0.7050
RF tuned @ 0.194 (Max F1)	0.8682	**0.3355**	**0.3627**	**0.3486**	0.7050

Key Drivers

- Margins & value: net_margin, margin_gross_pow_ele, margin_net_pow_ele.
- Usage behaviour: cons_12m and cons_last_month (recent vs longer-term usage).
- Lifecycle & pricing: months_activ/tenure and price-spread/volatility signals.

Recommendations

- Use churn probabilities to rank customers; avoid relying on a 0.5 threshold.
- Adopt tuned threshold 0.194 (Precision 0.34, Recall 0.36) and validate via retention A/B testing.
- Prioritise renewal-window and price-sensitive segments to improve retention ROI.

Note: Dataset not included in repository. Results shown for portfolio case study only.

Owen Nda Diche | Data Science Portfolio