Persistancy Of Drugs

For Data Glaciers
Internship - LISUM43

2025 May
Istiyak Shaikh

Overview

ABC Pharma wants to understand patient drug persistency—whether patients continue to take their medications as prescribed by physicians. The goal is to build a machine learning model that predicts persistency using patient demographics, clinical history, risk factors, and treatment behavior. Automating this process will help physicians and the pharma company improve adherence strategies and personalize patient interventions.

Predicting Patient Drug Persistency for ABC Pharma

Business Goal: Identify patients at risk of non-persistence to improve adherence and personalize interventions.

Value Proposition:

- Improve health outcomes
- Reduce cost from non-adherence
- Enable proactive outreach by physicians

Dataset Summary

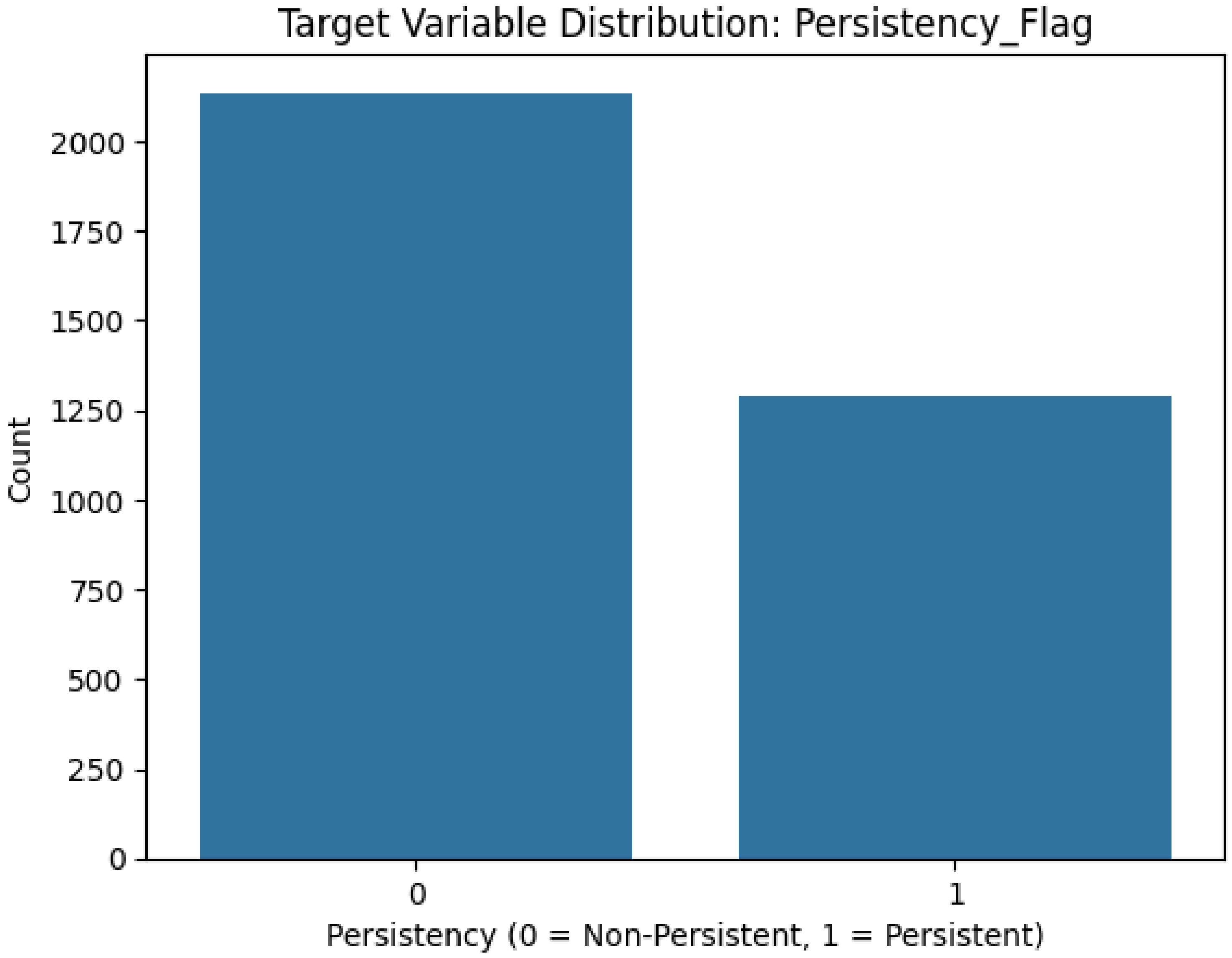
Metric	Value
Total Rows (Patients)	3424 (example)
Total Columns	25+ features
Target Variable	Persistency_Flag
Numeric Features	e.g., Dexa_Freq_During_Rx, Count_Of_Risks
Categorical Features	Gender, Age_Bucket, Race, Region, etc.
Risk Factor Flags	21 binary features like Risk_Estrogen_Deficiency

Target Variable Distribution

Persistency Status:

Persistent: 38%

Non-Persistent: 62%



🔍 Observation: The data is imbalanced toward non-persistent patients.

📈 Implication: Business should focus interventions on this larger group.

Patient Demographics

Age Buckets: Persistency slightly higher in patients >75 (1400+)

Gender: Large variation (Only 194 male data)

Race & Ethnicity: Certain groups (e.g., Hispanic) show lower persistency (98)

Region: Differences in adherence patterns across regions

Recommendation ⟶ Consider regional patient support strategies.

Risk Factor Impact on Persistency

- **Estrogen Deficiency** → Higher non-persistence
- **Low Calcium Intake** → Imbalanced (‘Y’ - 46 , ‘N’ - 3382)
- **Recurring Falls** → Correlates with lower adherence (‘Y’ - 69 , ‘N’ - 3315)

📌 **Suggestion:** Integrate clinical risk profiles into outreach targeting.

Treatment Behavior Insights

- **Dexa_Freq_During_Rx:** Strong right-skew with outliers
- **Specialist vs. Non-Specialist:** Patients under specialists tend to be more adherent
- **Treatment Bucket:** Certain treatment types show lower persistency

💡 Insight: Treatment behavior is a strong signal for adherence prediction.

Outliers & Data Skew

- **Outliers:** Detected in Dexa_Freq_During_Rx, winsorized or log-transformed
- **Skewed Features:** Addressed via transformations
- **Missing Values:** None (dataset is clean)

✅ **Confidence:** Data is model-ready after transformation and encoding

Technical Slide – Model Recommendation

- Recommended Model: Gradient Boosting + SHAP
- Best performance on hold-out set
- SHAP values used for feature-level interpretability
- Aligns with business need for actionable insights

Technical Slide – Model Recommendation

Model Type	Model	Notes
Baseline	Logistic Regression	Simple, interpretable, performs well with balanced preprocessing
Tree-Based	Random Forest	Good performance + feature importance, explainable
Boosting	Gradient Boosting (GBM)	Best predictive performance, handles imbalance well
Interpretable ML	Explainable Boosting Machine	Trade-off between accuracy & explainability

Our Team

Istiyak Shaikh

Data Scientist

Thank you