Healthcare Data Analysis

Predicting Drug Persistency

Group Name: The Closer
Member 1 Name: Rongala Sreedhar
Member 1 Email: rongalasreedhar@gmail.com
Member 1 Country: Italy
Member 1 College/Company: University of Naples Federico II
Member 1 Specialization: Data Science

Project Overview

Objective: To analyze factors affecting Drug Persistency and build a predictive model.

- Understanding patient demographics and comorbidities.
- Identifying key risk factors.
- Developing a Machine Learning model to predict persistency.
- Creating an interactive dashboard for stakeholders.

EDA: Patient Demographics

Key Insight: Gender distribution shows slight variation.

EDA: Regional Trends

Key Insight: Significant regional disparities in persistency rates.

EDA: Top Risk Factors

Key Insight: Comorbidities heavily influence persistency.

Modeling Strategy

We explored three families of models to balance accuracy and interpretability:

- Linear Model: Logistic Regression (Baseline, High Interpretability).
- Ensemble Model: Random Forest (Feature Importance).
- Boosting Model: XGBoost (High Accuracy).

Model Performance

- Logistic Regression: ~80% Accuracy | AUC: 0.82
- Random Forest: ~86% Accuracy | AUC: 0.88
- XGBoost: ~88% Accuracy | AUC: 0.90
-
- Recommendation: XGBoost selected for production due to highest predictive performance.

Model Explainability

Even with the Black Box model (XGBoost), we can use Feature Importance to understand drivers:

- Top Factor: Dexa_Freq_During_Rx
- Secondary Factor: Comorbidity_Index
- Tertiary Factor: Region

Interactive Dashboard

A Streamlit dashboard was developed to allow stakeholders to:

- Visualize dataset statistics.
- Input patient details.
- Get real-time persistency predictions.

Conclusion & Next Steps

- Data quality improved through preprocessing.
- Robust predictive model achieved (AUC ~0.90).
- Actionable insights on regional and risk-based targeting.
- Next Steps: Deploy dashboard and monitor model drift.

Code & Deliverables

Full Source Code and Report available at:

https://github.com/sreedharsiddhu/Data-Glacier