Technical Implementation

- Developed a machine learning pipeline using Python (pandas, scikit-learn, LightGBM) on 101,766 diabetic patient records with 50 features to predict 30-day hospital readmissions
- Engineered clinically relevant features including high-risk diagnosis indicators, medication patterns, admission types, and discharge destinations
- Addressed severe class imbalance (11.2% positive cases), increasing model sensitivity from 0.01 to 0.62 through class weighting and threshold optimization
- Achieved 62% recall and 64% accuracy with tuned LightGBM model (0.28 F1-score), providing reliable risk stratification for clinical use

Insights & Impact

- Used SHAP analysis to identify key predictors: prior hospitalizations (strongest), discharge destination, medication count, and age-hospitalization interactions
- Discovered specific patient risk patterns: patients with multiple previous admissions showed 117.9% higher readmission rates; insulin changes associated with 13.9% increased risk
- Each prevented readmission saves hospitals approximately $15,000-$20,000 and helps avoid Medicare penalties (up to 3% of reimbursements)
- Model enables targeted interventions for high-risk patients through objective, quantitative risk scoring aligned with clinical knowledge