**Project Proposal**

**Predicting Hospital Readmissions Using Machine Learning**

**Objective**

This project aims to build and evaluate machine learning models that predict patient readmissions within 30 days of discharge. By analyzing structured hospital data, the study will identify key risk factors and develop interpretable prediction models that support better patient outcomes and reduce hospital costs.

**Dataset**

The **Diabetes 130-US hospitals dataset (1999–2008)** will be used, which includes over 100,000 hospital admissions and features such as patient demographics, diagnoses, length of stay, lab tests, and discharge outcomes. The target variable is whether a patient is readmitted within 30 days.

**Methods**

* **Data preprocessing:** handle missing values, encode categorical variables, and address class imbalance.
* **Machine learning models:** Logistic Regression, Decision Trees, Random Forest, Gradient Boosting (XGBoost/LightGBM).
* **Model explainability:** SHAP or LIME to highlight important features influencing predictions.

**Evaluation Metrics**

* Accuracy (baseline comparison)
* ROC-AUC
* Precision, Recall, and F1-score
* Precision-Recall AUC
* Fairness metrics across demographic subgroups

**Expected Outcome**

The project will deliver a comparative analysis of machine learning models for predicting hospital readmission. Outcomes will include predictive performance benchmarks, feature importance insights, and interpretable explanations. The findings aim to demonstrate the potential of machine learning in improving hospital decision-making while highlighting considerations for fairness and healthcare equity.