

# PREDICTIVE MODELING FOR DIABETES-RELATED READMISSIONS IN U.S HOSPITALS

---

IE7300 Project Group 3

---

## Introduction

In the dynamic field of healthcare analytics and predictive modeling, we address the pressing issue of diabetes-related hospital readmissions. These readmissions pose a significant challenge to healthcare due to their financial burden and impact on patient well-being. Leveraging the "Diabetes 130-US hospitals for years 1999-2008" dataset from the UCI Machine Learning Repository, our project's aim is to develop a predictive model. Diabetes, a prevalent chronic condition, demands continuous care, and by understanding the factors contributing to readmissions, we can empower healthcare providers to take proactive steps to enhance care quality and mitigate the financial implications.

## Dataset Overview

The [dataset](#), obtained from UCI's repository, consists of information from 130 U.S. hospitals spanning the years 1999 to 2008. It provides a comprehensive set of attributes, including patient demographics, clinical and treatment-related factors, and hospital operational data. The primary dataset comprises over 100,000 instances, each representing a hospital admission, and 47 features. The key information we have at our disposal includes patient demographics, admission sources, clinical attributes (e.g., diagnoses, procedures), medication information, and various hospital-related factors.

## Project Goals

The primary objectives of this project are as follows:

### Predict Readmission Risk:

Develop a predictive model to estimate the risk of diabetes-related readmissions within 30 days of discharge from an initial hospitalization. This predictive model will help healthcare providers identify patients at high risk and intervene to reduce readmissions.

### Feature Selection and Engineering:

Identify the most relevant features for readmission prediction and potentially create new features or feature transformations to enhance model performance.

## Model Evaluation:

Employ appropriate machine learning algorithms to build and evaluate predictive models. Model evaluation metrics such as accuracy, precision, recall, F1-score, and ROC AUC will be utilized to assess model performance.

## Interpretability:

Ensure that the model provides interpretable insights into the factors contributing to readmissions, enabling healthcare practitioners to take actionable steps to reduce readmission rates.

# Project Plan

## Data Preprocessing:

- **Data Cleaning:** Address missing values and inconsistencies in the dataset
- **Feature Selection:** Identify relevant features for the predictive model
- **Data Transformation:** Encode categorical variables and standardize numerical variables

## Exploratory Data Analysis (EDA):

- Perform in-depth EDA to gain insights into the dataset, understand the distribution of key variables, and discover potential correlations.

## Model Development:

- Split the dataset into training and testing sets
- Develop predictive models using various algorithms
- Tune hyperparameters to optimize model performance

## Model Evaluation:

- Evaluate models using metrics like accuracy, precision, recall, F1-score, and ROC AUC.
- Perform cross-validation to ensure model generalization.

# Problem Statement

This project aims to develop a predictive model for diabetes-related hospital readmissions in U.S. hospitals. By leveraging the UCI dataset and following a structured plan, we will create a valuable tool that can assist healthcare providers in identifying high-risk patients and taking proactive measures to reduce readmissions. The ultimate goal is to improve patient outcomes, reduce healthcare costs, and enhance the quality of care in diabetes management.