

Joseph Bailey

Kristen Lowe

Zan Merrill

Varsha Ramesh

Nisha Sapkota

## **AML Group 19 Project Proposal**

### **Predicting Hospital Readmission and Cost Risk Using Machine Learning**

#### **Objective**

In this project, we will be using machine learning to develop a predictive model for hospital readmission risk for patients with diabetes. Treatment success is defined by whether or not the patient is readmitted to the hospital, either within 30 days or after 30 days.

#### **Scope**

Our analysis will also explore cost risk analysis by identifying high-utilization patients, fairness across subpopulations (e.g. race, age, gender) to evaluate if risk predictions introduce bias, and feature interpretability to make results actionable for hospitals and healthcare administrators.

#### **Problem Statement**

Hospital readmissions among diabetic patients represents a major challenge for healthcare systems, leading to higher costs and worse patient outcomes. Despite the availability of various treatment and care plans, many patients with diabetes are readmitted (some within 30 days of being discharged) due to the wrong treatment strategies. Our project aims to identify the most effective treatment approaches and management practices for hospitalized diabetic patients to minimize the likelihood of readmission. By analyzing demographic data, preventative and therapeutic prescriptions, and clinical tests, our goal is to develop recommendations that improve long-term patient health and reduce total cost of care.

#### **Business Use Case**

Preventing avoidable readmissions is valuable for:

Stakeholder	Value Provided
Hospitals	Reduce penalties from CMS, improve quality metrics
Insurance Providers	Reduce total cost of care
Care Management Teams	Identify high-risk patients for intervention programs
Patients	Improve care outcomes

This system could be integrated into hospital care management workflows to flag high-risk patients before discharge and guide follow-up actions (e.g., telehealth check-ins, medication counseling).

## Dataset Description

### **Source:**

The dataset that we will use is provided by UC Irvine and represents 10 years (1999–2008) of care across 130 hospitals in the United States.

### **Size:**

The data includes 101,766 admissions, 47 features, and a mix of categorical and integer feature types.

### **Missing values:**

There are also some missing values denoted by question marks that we will have to pay attention to.

### **Inclusion criteria:**

An admission is included if it satisfies five criteria.

- The patient must have been admitted,
- diabetes was entered into the system as a diagnosis at some point,
- the length of stay was at least 1 full day and no more than 14 days,
- lab tests were performed,
- and some medication was prescribed.

### **Feature groups:**

The features include demographic questions such as race, gender, age, and weight; admission and discharge questions; and clinical questions like what tests were performed and what medicines were administered.

### **Target classes:**

Ultimately this dataset classifies each patient as being readmitted within 30 days, readmitted after more than 30 days, or not readmitted to the hospital at all.

### **Proposed Approaches**

For our methodology, we will divide our workflow into three stages: preprocessing, modeling, and evaluation. During preprocessing, we will address any missing values in the dataset and split the data into training and test sets to validate our results. For modeling, we will use logistic regression to predict whether patients are readmitted, applying an L1 (Lasso) penalty to eliminate less effective factors, particularly medications that do not help prevent readmission. We will also test other models against this one and compare results. Finally, in the evaluation stage, we will focus on maximizing the accuracy of predicting patients who will not require readmission based on key factors. These results will help identify the most effective clinical tests and medications for reducing hospital readmissions among diabetic patients.

### **References**

<https://archive.ics.uci.edu/dataset/296/diabetes+130-us+hospitals+for+years+1999-2008>

<https://onlinelibrary.wiley.com/doi/10.1155/2014/781670>

<https://www.cms.gov/medicare/quality/value-based-programs/hospital-readmissions>