## 🏥 Hospital Readmission Prediction (30-Day Readmission Risk)

Machine Learning project that predicts whether a patient will be **readmitted within 30 days** after hospital discharge.
This system provides an **early warning risk score** to help hospitals improve patient safety, reduce costs, and optimize capacity.

---

### 📌 Project Objective

Hospital readmissions within 30 days are a major problem for the healthcare system:

- Increased hospital costs
- Overload on doctors and nurses
- High risk for chronic and elderly patients
- Negative impact on hospital quality metrics (Readmission Rate)

This project aims to build a **predictive model** that identifies high-risk patients *before* discharge, enabling preventive interventions.

---

### 🩺 Problem Description

Predicting readmission risk is extremely challenging for clinicians because:

- Patient medical history is complex
- Chronic diseases vary widely
- Home recovery can't be monitored
- Social & environmental factors affect outcomes

A data-driven ML model can support decision-making by providing **objective risk predictions**.

---

### 🎯 Success Metrics

The model performance was evaluated using:

- **ROC-AUC ≥ 0.70**
- **PR-AUC** (ideal for imbalanced datasets)
- Minimizing **false negatives**
- Balanced performance after SMOTE
- SHAP interpretability for explainable AI

---

### 📊 Dataset

**Source:** UCI Machine Learning Repository — Diabetic Hospital Readmission Dataset
**Size:** ~100,000 patient records

Key features include:

- Age group
- Gender
- Number of previous inpatient visits
- Emergency visits
- Number of laboratory procedures
- Number of medications
- Insurance & discharge type
- **Target:** readmitted (NO, >30, <30)

The target was converted to a **binary classification**:

- **1:** Readmitted within 30 days
- **0:** Not readmitted

---

### 🔧 Technical Approach

#### 1️⃣ Data Preparation & EDA

- Missing value treatment
- Outlier handling (IQR & winsorization)
- Correlation analysis
- Categorical encoding
- Feature engineering

#### 2️⃣ Imbalanced Data Solution

Dataset distribution:

- **11% positive (readmitted)**
- **89% negative**

Solution: **SMOTE Oversampling**

#### 3️⃣ Modeling

Main model: **XGBoost Classifier**

Best performance:

- **ROC-AUC: 0.74**

- **PR-AUC: 0.32**

### 4️⃣ Explainable AI (SHAP)

Used SHAP to explain model outputs:

- SHAP Summary Plot
- Feature Importance
- Doctor-friendly interpretation

This creates a transparent, trustworthy model.

---

### ⚙️ FastAPI Service (Backend)

The trained ML model is served via a REST API built with **FastAPI**.

**Run API:**

uvicorn app:app --reload

Swagger documentation:
👉 http://127.0.0.1:8000/docs

---

### 🖥️ Streamlit User Interface

A clean and user-friendly web interface for hospital staff to enter patient details and obtain risk predictions.

**Run Streamlit:**

streamlit run ui.py

---

### 📂 Project Structure

project/

│

├── data/

│ └── diabetic_data.csv

│

├── models/

│ └── readmission_xgb.pkl

│

├── app.py

├── main.py

├── ui.py

├── shap_summary.png

└── README.md

---

## 🚀 Installation

**1. Create a virtual environment**

python -m venv .venv

**2. Activate environment**

.venv\Scripts\activate

**3. Install dependencies**

pip install -r requirements.txt

---

## 🧩 Challenges & Solutions

| Challenge | Solution |
|---|---|
| Many missing values | Median imputation & dropping invalid rows |
| Highly imbalanced dataset | SMOTE oversampling |
| Numerous categorical features | One-hot & target encoding |
| Need for interpretability | SHAP explainability |
| API + UI deployment | FastAPI + Streamlit integration |

---

## 🏁 Conclusion

This project provides a **complete early warning decision support system** for healthcare.
It allows hospitals to:

- Identify high-risk patients
- Reduce readmission rates
- Improve quality metrics
- Enhance patient safety

With ML, API, UI, and explainability integrated end-to-end, this system can easily be extended for real-world use.

---

## 👩‍💻 Developer

**Busenur Durak**
Management Information Systems
Izmir Bakırçay University
AI / Data Analytics