**🎯 Objective**

To predict **patient outcomes** (e.g., readmission risk, recovery probability, or discharge outcome) using hospital data and machine learning — enabling healthcare providers to make data-driven decisions and improve patient care.

**🧩 Key Components**

| **Component** | **Description** |
| --- | --- |
| **Data Source** | Electronic Health Records (EHR) containing demographic, clinical, and discharge details. |
| **Features Used** | age, num\_procedures, days\_in\_hospital, comorbidity\_score, gender, primary\_diagnosis, discharge\_to. |
| **Target Variable** | Patient outcome (readmitted ). |
| **Model** | XGBoostClassifier trained with categorical encoding (OneHotEncoder + ColumnTransformer). |
| **Preprocessing** | Numeric features: passed through.  Categorical features: one-hot encoded. |
| **Deployment** | Model saved as .pkl using joblib, hosted on cloud storage or server. |
| **Dashboard** | Streamlit web interface for interactive user input and live predictions. |

**⚙️ Pipeline Workflow**

**Step 1 – Data Preparation**

* Data collected from hospital records
* Missing values handled, categorical data encoded
* Split into training and test sets

**Step 2 – Model Training**

* XGBoost Classifier trained with cross-validation
* Evaluation metrics: Accuracy, F1-Score, Confusion Matrix

**Step 3 – Model Saving**

* Trained model and preprocessing pipeline saved via joblib

**Step 4 – Deployment**

* Files (xgb\_model.pkl, preprocessor.pkl) uploaded to cloud (e.g., AWS S3, GCP Storage, Azure Blob, or GitHub repo)

**Step 5 – Streamlit Dashboard**

* User inputs patient details
* Input transformed using the same preprocessor
* Model predicts outcome in real time
* Dashboard displays results and probabilities

**☁️ End-to-End Architecture Diagram**

Here’s a clean conceptual diagram of the entire system flow 👇

┌────────────────────────────┐

│ Hospital Data │

│ (EHR: Demographics, Labs) │

└──────────────┬──────────────┘

│

▼

┌────────────────────────────┐

│ Data Preprocessing │

│ • Cleaning & Encoding │

│ • Feature Engineering │

└──────────────┬──────────────┘

│

▼

┌────────────────────────────┐

│ Model Training (ML) │

│ XGBoost Classifier │

│ + ColumnTransformer(OHE) │

└──────────────┬──────────────┘

│

▼

┌────────────────────────────┐

│ Model Export (.pkl) │

│ + Preprocessor.pkl │

└──────────────┬──────────────┘

│

▼

┌──────────────────────────────────┐

│ Cloud Storage │

│ (AWS S3 / GCP / Azure / GitHub) │

└────────────────┬──────────────────┘

│

▼

┌────────────────────────────┐

│ Streamlit Dashboard │

│ • User inputs patient info │

│ • Applies preprocessing │

│ • Model predicts outcome │

│ • Displays risk/probability

└────────────────────────────┘

**📊 Tech Stack**

| **Layer** | **Tools / Libraries** |
| --- | --- |
| **Data Handling** | pandas, numpy |
| **Modeling** | xgboost, scikit-learn |
| **Persistence** | joblib |
| **Dashboard** | streamlit |
| **Cloud / Hosting** | AWS / GCP / Azure / GitHub Pages / Streamlit Cloud |

**📈 Expected Outcomes**

* Improved prediction of patient recovery or readmission.
* Reduced manual decision errors through data-driven insights.
* Deployable end-to-end ML workflow from data → model → dashboard.