Hospital Readmission Prediction | AI Project '25

Objective:

Develop a machine learning system to predict whether a diabetic patient will be readmitted to the hospital within 30 days based on their medical history, treatment information, and demographic details.

Dataset Overview:

The dataset contains 101,766 records and 50 features.

- File: diabetic_data.csv
- Target: readmitted (Values: NO, >30, <30)
- Suggested Type: Multiclass or Binary classification (Readmitted <30 days = 1, others = 0)
- Data is not pre-split. Recommended split: 80% training / 20% testing using **stratified** sampling

ID Mapping File Explanation:

Some features use numerical codes which are not human-readable. Use 'IDs_mapping.csv' to decode them.

- Columns using mappings: admission_type_id, discharge_disposition_id, admission_source_id
- Example from mapping file:

admission_type_id	description
1	Emergency
2	Urgent
3	Elective
4	Newborn
5	Not Available

Requirements:

1. **Data Preprocessing**

- Handle missing values and decode IDs using mapping file
- Encode categorical features using label/one-hot encoding
- Normalize or standardize numerical columns
- Remove or handle outliers

2. **Exploratory Data Analysis (EDA)**

- Feature distribution plots and correlation analysis
- Investigate relationships between readmission and diagnoses or medications

3. **Model Development**

- Train and compare at least 4 ML algorithms (e.g., Logistic Regression, SVM, Random Forest, XGBoost)
- Use stratified split: train.csv for training, test.csv for evaluation
- Use metrics: accuracy, precision, recall, F1-score, confusion matrix, ROC-AUC

4. **GUI Application**

- Build using Tkinter or PyQt
- User inputs patient info → get prediction
- Show prediction with confidence score
- Optional: Visual explanations like feature importance

5. **Final Reporting (PDF)**

- Include all preprocessing and model steps
- Present EDA insights and feature impact
- Compare models and tuning strategies
- Final summary of outcomes and insights