Hospital Patient Encounter Classification

Author: Bangaly Sano

Date: May 2025

1. Introduction

This project focuses on classifying hospital patient encounters using machine learning. The goal is to predict the type of hospital encounter (e.g., inpatient, outpatient, emergency) based on electronic health record (EHR) features.

2. Dataset Overview

The dataset contains patient records with fields such as REASONCODE, ENCOUNTERTYPE, PROVIDERID, and others. It includes a mix of categorical and numerical variables and requires preprocessing before modeling.

3. Data Preprocessing

Key preprocessing steps included:

Handling missing values

Converting date fields to datetime format

Creating a new feature for encounter duration

Encoding categorical variables

Scaling numerical data using StandardScaler

4. Model Building

A Random Forest Classifier was used for modeling. The dataset was split into training and testing sets, and the classifier was trained on the processed features to predict the ENCOUNTERCLASS.

5. Model Evaluation

The model achieved 100% accuracy on the test set. The classification report showed perfect precision, recall, and F1-score across all classes. A confusion matrix was also generated to visualize performance.

6. Visualizations

Two main plots were generated:

Confusion Matrix

Feature Importance Bar Chart

These plots help interpret the model's effectiveness and understand which features contributed most.

7. Conclusion

This project demonstrates the effectiveness of Random Forest for healthcare classification tasks. With clean data and strong feature engineering, machine learning models can predict encounter types with high accuracy.

8. Tools Used

Python

Jupyter Notebook

Pandas, NumPy

Scikit-learn

Matplotlib, Seaborn