Predicting hospital readmissions for diabetes patients

A Complete End to End Healthcare Analytics Project

Duration: 2-3 Hours | Recorded Lectures Available

@ashoks773

@ashoks773

@sharma-ak

Instructor: Dr. Ashok Kumar Sharma
Bioinformatics Scientist (Senior Data Analyst)
Cedars Sinai Medical Center, Los Angeles, CA

Personal webpage:
https://ashoks773.github.io

Follow: @datawithashok   YouTube

Learnings:

- Acquire and validate healthcare data

- Process messy real-world datasets

- Build predictive models

- Create professional visualizations

- Document everything on GitHub

Segment 1: Introduction & Data Acquisition

- Project overview and objectives

- Data source introduction (UCI)

- Environment setup (Python)

- Data download and initial loading

PROJECT OVERVIEW
Focus: End-to-end healthcare analytics project

Problem: Predicting 30-day hospital readmissions to reduce costs and improve patient outcomes

Tool: Python (exclusively)

Dataset: Diabetes 130-US Hospitals (1999-2008) - UCI Machine Learning Repository

Segment 2: Data Preprocessing & Cleaning

- Data quality assessment

- Missing value treatment

- Feature engineering

- Data transformation

Segment 3: Exploratory Data Analysis

- Univariate analysis

- Bivariate analysis

- Key insights identification

- Visualization creation

Segment 4: Exploratory Data Analysis

- Model building

- Model evaluation

- Feature Importance analysis

Segment 5: Reporting and Documentation

- Executive summary creation

- GitHub repository setup ➡️

- Documentation best practices

```
## 📁 Project Structure
```

healthcare-readmission-prediction/
│
├── diabetic_data.csv   # Raw dataset (download separately)
├── diabetic_data_cleaned.csv   # Preprocessed dataset
├── healthcare_analytics.py   # Main analysis script
├── EXECUTIVE_SUMMARY.txt   # Executive summary report
├── README.md   # This file
│
├── visualizations/   # All generated visualizations
│   ├── 01_target_distribution.png
│   ├── 02_numeric_distributions.png
│   ├── 03_categorical_distributions.png
│   ├── 04_numeric_vs_readmission.png
│   ├── 05_categorical_vs_readmission.png
│   ├── 06_correlation_matrix.png
│   ├── 07_model_comparison.png
│   ├── 08_confusion_matrices.png
│   ├── 09_feature_importance.png
│   ├── 10_probability_distribution.png
│   ├── 11_risk_stratification.png
│   └── 12_FINAL_DASHBOARD.png
│
└── requirements.txt   # Python dependencies
```