Hospital Readmission Analysis & Prediction

Objective:

To identify key factors influencing patient readmission and develop a predictive model to forecast the likelihood of readmission, enabling improved healthcare outcomes and cost management.

Tools and Technologies Used:

- Python (pandas, numpy, scikit-learn, matplotlib, seaborn, XGBoost)
- Jupyter Notebook
- HTML export for report presentation

Overview of Work Performed:

Data Understanding and Cleaning:

- Loaded and explored the hospital readmission dataset.
- Checked for null values and outliers.
- Performed data cleaning by removing or imputing missing values.
- Converted categorical variables into numerical using encoding techniques like One-Hot Encoding and Label Encoding.

Exploratory Data Analysis (EDA):

- Performed statistical analysis to understand data distribution and correlation between variables.
- Created visualizations (bar charts, heatmaps, histograms) to analyze patient demographics, treatment types, and readmission rates.
- Identified key trends such as age groups with high readmission rates, commonly administered medications, and diagnoses linked to frequent readmissions.

Feature Engineering:

- Created new features such as 'Total Visits', 'Total Procedures', and 'Medication Changes'.
- Transformed features for better model performance (e.g., binning age into ranges).

Model Building and Evaluation:

- Split the dataset into training and testing subsets.
- Implemented multiple classification models: Logistic Regression, Decision Tree, Random Forest, and XGBoost.
- Evaluated models using accuracy, precision, recall, F1-score, and ROC-AUC.
- XGBoost was found to be the most effective model with high accuracy and interpretability.

Model Interpretation and Insights:

- Used SHAP and feature importance plots to interpret model results.
- Identified top predictors of readmission including number of inpatient visits, change of medication, discharge disposition, and admission type.

Deployment Readiness:

- Prepared the model for integration into a clinical decision support system.
- Documented the preprocessing pipeline and model deployment steps.

Conclusion:

The analysis successfully highlighted critical factors influencing hospital readmissions and developed a robust machine learning model for prediction. The insights generated can help healthcare providers implement targeted interventions and reduce unnecessary readmissions.

.