Diabetes Readmission Prediction Project

Author: IBRAHIM AKINTUNDE AKINYERA
Email: Akinyeraakintunde@gmail.com
GitHub: https://github.com/akinyeraakintunde/Diabetes-Readmission
Portfolio: https://akinyeraakintunde.github.io/Diabetes-Readmission/

Abstract

This project predicts hospital readmissions for diabetic patients using machine learning on over 100,000
It applies data cleaning, feature engineering, and models such as Random Forest and XGBoost to iden
The results provide actionable insights and demonstrate practical machine learning skills..

Introduction

Hospital readmissions are costly and impact patient health.
This project analyzes diabetic patient data to predict readmissions and identify key factors influencing r

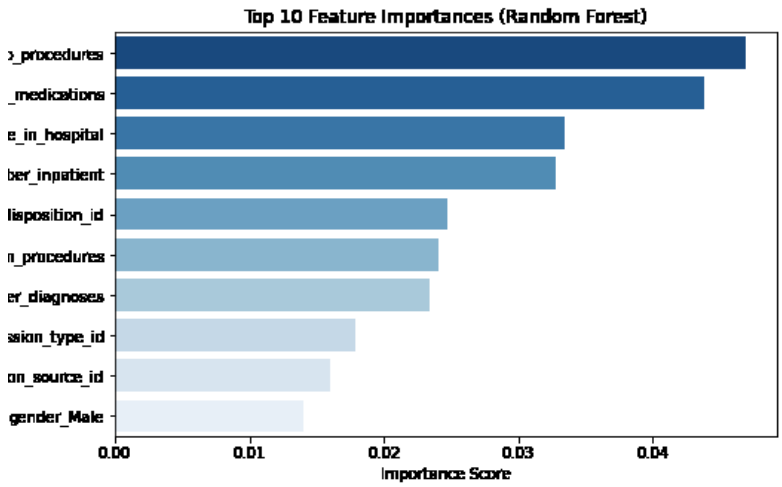
Dataset Description

Dataset Source: Kaggle Diabetes Healthcare Dataset.
Records: 101,766.
Features: 50+.
Key columns include age, inpatient visits, medications, admission type, readmission status..

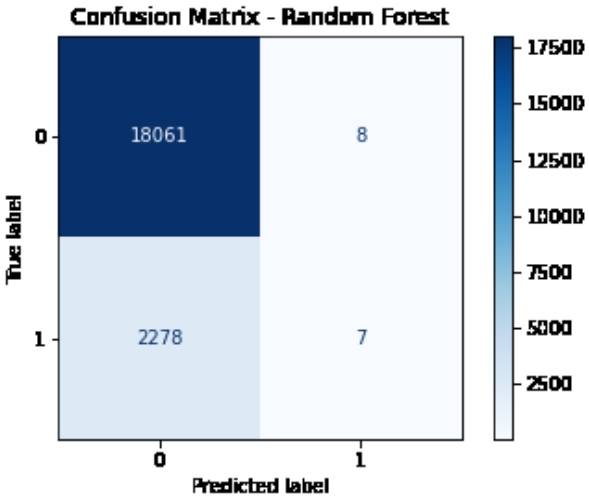
Methods

Data Cleaning: Handle missing values, remove duplicates.
Feature Engineering: Encode categorical variables, create new features.
Models Used: Random Forest, XGBoost, Logistic Regression.
Evaluation Metrics: Accuracy, Confusion Matrix, ROC Curve..

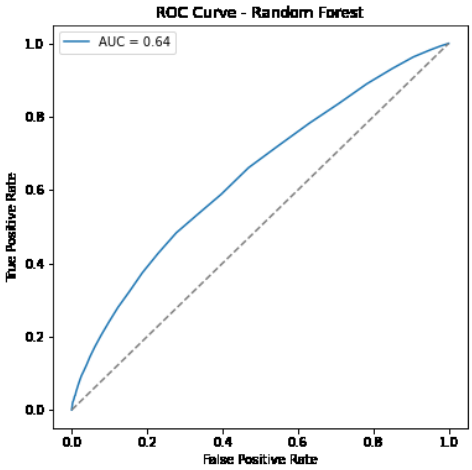
Feature Importance



Confusion Matrix



ROC Curve



Discussion & Conclusion

Random Forest identified age, number of inpatient visits, and medications as the most important predic
Limitations include missing data and class imbalance.
This project demonstrates practical ML skills, provides actionable healthcare insights, and is portfolio-re

References

Kaggle Diabetes Healthcare Dataset: https://www.kaggle.com/datasets/uciml/diabetes-healthcare-datas
Python libraries: Pandas, Scikit-learn, Matplotlib, Seaborn..