

Machine Learning Engineer Capstone Project

Masinde Mtesigwa

2023-03-01

Contents

Healthcare	2
EHR/EMR	2
Definition	2
The dataset	2
Project Statement	2
Benchmark model	2
Metrics	2

Healthcare

EHR/EMR

Electronic Health records or Electronic Medical Records data is the data being collected when we see a doctor, pick up a prescription at the pharmacy, or even from a visit to the dentist.

This data is used for a variety of use-cases. From personalizing healthcare to discovering novel drugs and treatments to helping providers diagnose patients better and reduce medical errors.

Definition

The objective of the project is to diagnostically predict whether or not a patient has diabetes, based on certain diagnostic measurements included in the dataset.

The dataset

The dataset is taken from Kaggle

ref:<https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database>

Project Statement

Can you build a machine learning model to accurately predict whether or not the patients in the dataset have diabetes or not?

Benchmark model

The model going to be used in this project is Autogluon or Classification algorithm since it is a classification problem.

Metrics

A set of evaluation metrics

- The confusion matrix is a technique used for summarizing the performance of a classification algorithm i.e. it has binary outputs.
- Classification Report
- ROC - AUC ROC (Receiver Operating Characteristic) Curve tells us about how good the model can distinguish between two things (e.g If a patient has a disease or no). Better models can accurately distinguish between the two. Whereas, a poor model will have difficulties in distinguishing between the two