

Classificazione del diabete

Introduzione

Nel contesto dell'assistenza sanitaria preventiva e della gestione delle risorse cliniche, le aziende sanitarie e gli assicuratori hanno sempre più bisogno di strumenti predittivi affidabili per identificare precocemente i soggetti a rischio di sviluppare il diabete. La possibilità di classificare i pazienti in base alla probabilità di malattia consente alle compagnie assicurative di valutare con maggiore precisione il rischio sanitario individuale, personalizzare le polizze assicurative, definire premi più equi e contenere le spese mediche attraverso programmi di prevenzione mirati.

Nel quadro di questa ricerca, l'obiettivo principale è quello di determinare se un paziente presenta una diagnosi di diabete, sulla base di dati clinici e anamnestici raccolti.

Obiettivo

Sviluppare un sistema di classificazione binaria per la diagnosi del diabete. La classe target ('diabetes') indica la presenza (1) o l'assenza (0) della malattia.

Indicazioni

Il dataset è fornito in due file CSV (train_diabetes.csv e test_diabetes.csv) contenente sia dati clinici sia il target di classificazione. Il dataset di training può eventualmente essere suddiviso per un validation set. Per una corretta valutazione del modello, i dati di test non devono mai essere utilizzati nella fase di addestramento (training).

Attributi

gender: Sesso del paziente (categorico).

age: Età del paziente (numerico).

hypertension: Indica se il paziente soffre di ipertensione (0 = No, 1 = Sì).

heart_disease: Indica se il paziente soffre di malattie cardiache (0 = No, 1 = Sì).

smoking_history: Storia del fumo del paziente (categorico).

bmi: Indice di massa corporea (numerico).

HbA1c_level: Livello di emoglobina glicata (numerico).

blood_glucose_level: Livello di glucosio nel sangue (numerico).

diabetes: Variabile target: 1 = presenza di diabete, 0 = assenza.

Insulin_Sensitivity_Est: Stima della sensibilità all'insulina (numerico).

BMI_Glucose_Interaction: Interazione tra BMI e glucosio (numerico).

Troponina_T: Marcatore laboratoristico altamente specifico per il danno cardiaco acuto, usato per diagnosticare infarti miocardici.