

Estudantes: Carolina Ferreira, Gustavo William, Mateus Caçabuena e Vitória Gonzalez

volumetria: 101767

propósito: diabetes

características:

- pacientes diabéticos de um hospital
- controle de fluxo / atividade dos pacientes com diabetes

qualidade dos dados:

- há dados faltantes? sim, em grande quantidade. Principalmente nos atributos “weight”, “payer_code” e “medical_specialty”. Para corrigir, nós eliminaríamos os atributos “weight” e “payer_code” que estão quase inteiramente incompletos, e em “medical_specialty” substituiríamos as “?” por “unknown”. Além disso, excluiríamos o atributo “acetohexamide” pois todos os seus valores são iguais, de forma que não interfere na saída.
- Sim, pode ser utilizado para classificar quem tem diabetes. A classe é ter ou não ter diabetes. Não está balanceado pois tem muito mais pessoas com diabetes do que sem. A distribuição por classe é binária pois só temos duas saídas: ter ou não diabetes. Nós usaríamos estes atributos para classificação: “diabetesMed”, “insulin”, “age” e “admission_type”.