

MC536 - Etapa 2 do Trabalho

Victor Prado Palmerini 178061

RESUMO

O trabalho proposto tem como objetivo resumir, a partir do agrupamento e visualização, dados relacionados a diagnóstico de diabetes nos Estados Unidos. Os dados são mapeados por estados e as análises estatísticas terão então esse objetivo de extrair informações relevantes sobre o dataset que será utilizado, como médias, desvios, agrupamento e identificação de padrões.

LISTA DE REQUISITOS

1. Pacientes devem ser únicos
2. Diagnósticos devem ser únicos para cada paciente
3. Pacientes devem informar idade e número de gravidez
4. Diagnósticos devem conter os seguintes dados: taxa de glicose, pressão sanguínea, largura da pele, taxa de insulina, BMI e função pedigree
5. Pacientes devem ser classificados se houve diagnóstico da diabetes ou não

FONTE DOS DADOS

Inicialmente a fonte de dados foi o site do Center of Disease Control and Prevention. Neste site é possível encontrar dados demográficos relacionados à incidência e distribuição de diabetes pelos estados dos Estados Unidos. Porém, os dados se mostraram bastante distribuídos e de difícil agrupamento e análise.

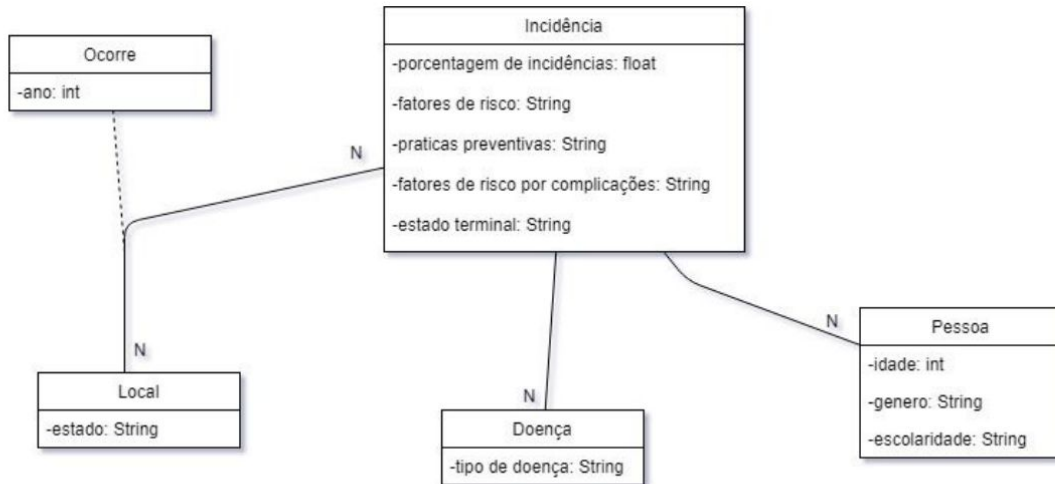
Center of Disease Control and Prevention: <https://gis.cdc.gov/grasp/diabetes/DiabetesAtlas.html>

A fonte de dados atual é o site Kaggle, que possui um dataset sobre diagnóstico de diabetes.

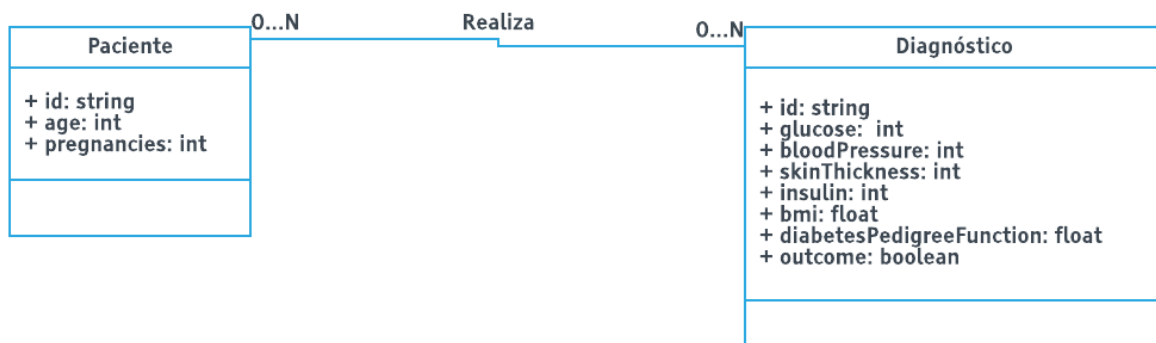
Kaggle: <https://www.kaggle.com/uciml/pima-indians-diabetes-database>

MODELAGEM CONCEITUAL - VERSÕES

1ª Versão



2ª Versão

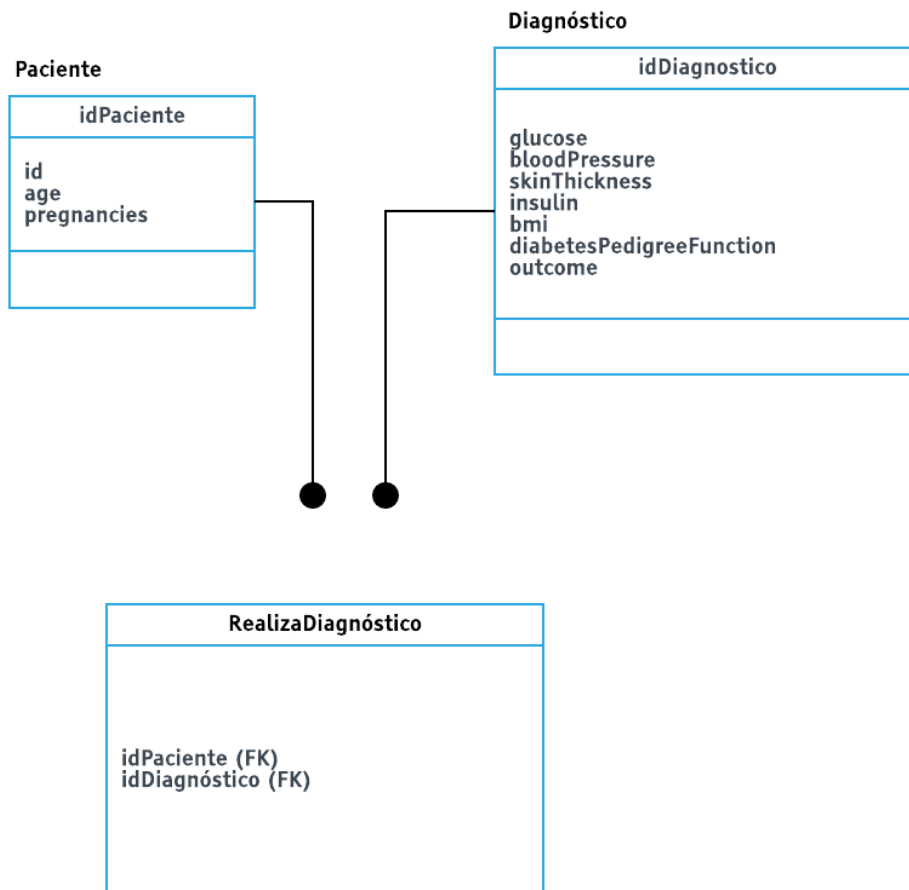


Modificações:

- 1) Alteração da fonte de dados acarretou numa mudança do modelo conceitual.

- 2) Ênfase maior no diagnóstico da diabetes e mais simples (menos entidades e relacionamentos).
- 3) Versão atual possui agora a entidade “Diagnóstico”.

MODELO LÓGICO



No modelo lógico ocorreu a adição da Tabela “RealizaDiagnóstico”, para representar a relação entre Paciente e Diagnóstico (usando chaves estrangeiras), além de definir as Tabelas “Paciente” e “Diagnóstico”, que representam as entidades “Paciente” e “Diagnóstico”, respectivamente.