

Grupo: Torresmo

Nome: Thamiris Florindo Coelho
Nome: Daniela Kemy Kagawa

RA: 187506
RA: 166293

Diagnóstico de doenças cardíacas

Resumo

O banco de dados conterá informações de diversos pacientes como sexo, idade, se fuma ou não, sintomas e resultados de exames, a fim de estudar a presença de doenças cardíacas e como elas se relacionam com esses dados. A partir disso, pretendemos utilizar aprendizado de máquina para dizer se um novo paciente possui ou não uma doença cardíaca, onde essas estão classificadas no intervalo de $[0,4]$, portanto o sistema deve retornar em qual faixa se encontra o paciente.

Requisitos

1. O sistema deverá ser capaz de armazenar dados de pacientes
2. O sistema deverá ser capaz de armazenar dados de exames de pacientes
3. O sistema deverá ser capaz de armazenar dados de sintomas de pacientes
4. O sistema deverá ser capaz de armazenar dados de diagnóstico de um paciente que está relacionado também com exames e sintomas
5. O sistema deverá determinar a propensão de determinado paciente, com seus respectivos exames e sintomas, possuir doenças cardíacas.

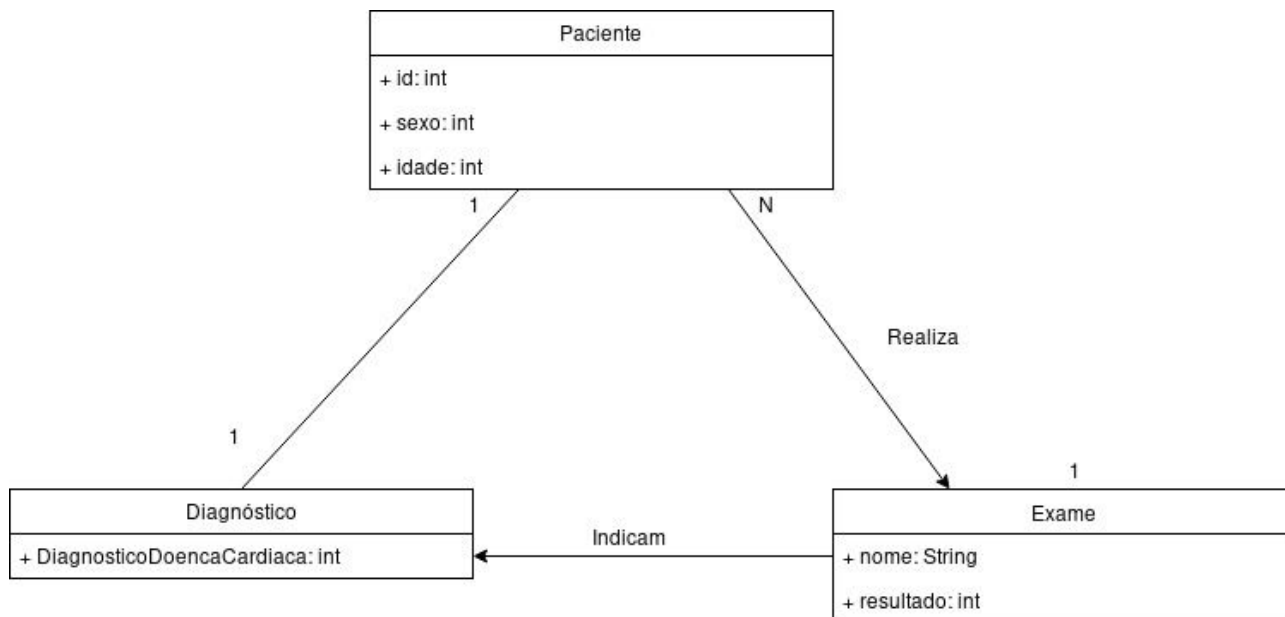
Site

O site utilizado para retirada da base de dados foi o Kaggle, ele é uma plataforma de modelagem preditiva e de competições analíticas em que estatísticos e mineradores de dados competem para produzir os melhores modelos para prever e descrever os conjuntos de dados enviados por empresas e usuários.

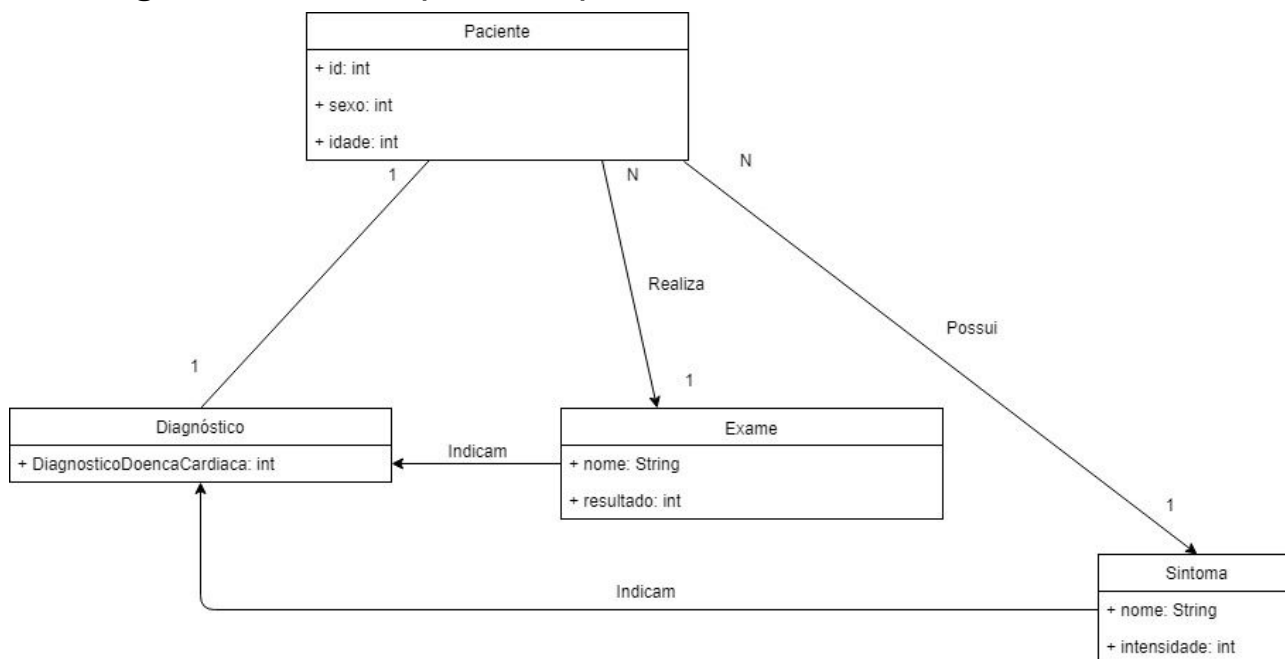
Base de Dados

A base de dados escolhida foi a base do desafio Heart Attack Prediction - Kaggle (<https://www.kaggle.com/imnikhilanand/heart-attack-prediction>). Essa base foi construída a partir de dados que foram coletados das seguintes localidades: *Cleveland Clinic Foundation*, *Hungarian Institute of Cardiology of Budapest*, *V.A. Medical Center of Long Beach* e *University Hospital of Zurich*. Originalmente a base continha setenta e seis atributos, mas os dados já foram processados e atualmente possui quatorze atributos e todos estão representados numericamente.

Modelagem conceitual (versão 1)



Modelagem conceitual (versão 2)



Alterações feitas da versão 1 para a versão 2

A alteração feita da Versão 1 para a versão 2 do modelo conceitual foi a adição a classe “Sintoma” com os atributos “nome” (do tipo String) e “intensidade” (do tipo inteiro), pois nem todos os dados apresentados na base de dados se encaixam na categoria de “Exame”. A classe “Sintoma”, assim como a classe “Exame”, se relaciona com a classe “Diagnóstico” e o relacionamento é de indicação (sintomas indicam um diagnóstico).

Modelo lógico

Paciente(idPaciente, sexo, idade)

Exame(idPaciente, nome, resultado)

□ idPaciente é CHE para Paciente

Sintoma(idPaciente, nome, intensidade)

□ idPaciente é CHE para Paciente

Diagnostico(idPaciente, diagnostico)

□ idPaciente é CHE para Paciente

Comentários a respeito do modelo lógico

Para adequar o modelo conceitual ao modelo lógico, foram introduzidas chaves estrangeiras (“idPaciente”) em Exame, Sintoma e Diagnostico, a fim de tornar possível o relacionamento entre as diferentes tabelas. Além disso, foram feitas pequenas modificações nos nomes dos atributos para facilitar futuras consultas realizadas a partir dessas tabelas.