

MC536 - Banco de Dados

Trabalho Etapa 2

Nome: Murilo Guidetti Andrietta

Ra: 147472

Nome: Naomi Takemoto

Ra: 184849

Primeira Etapa:

O primeiro modelo conceitual desenvolvido focava em uma organização de dados que estivesse próxima daquilo que se espera para uma implementação do algoritmo de machine learning. Nesse caso, ocorreu um erro de modelagem ao fazer o modelo conceitual se preocupando com características que deveriam estar no modelo lógico.

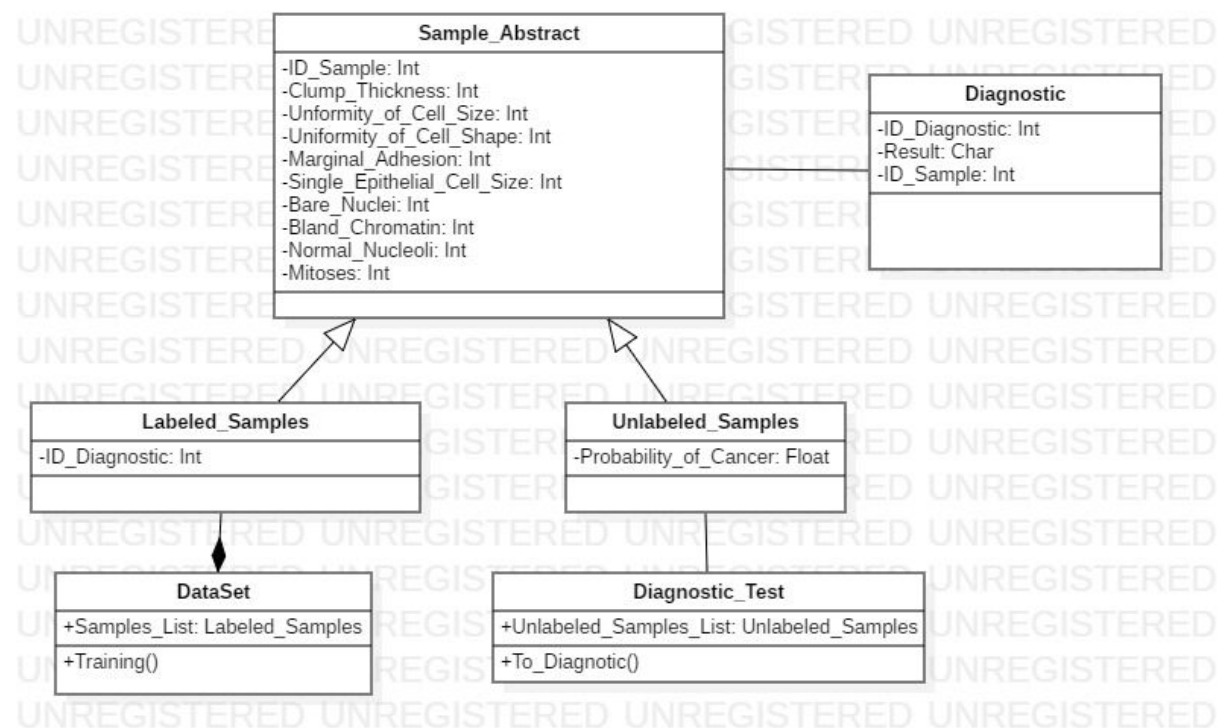


figura 1: modelagem conceitual da primeira parte do trabalho (apresentação)

Segunda Etapa:

Para o desenvolvimento da etapa 2, levou-se em consideração funcionalidades que uma aplicação de diagnóstico de câncer de mama deveria ter. Nesse caso, o sistema deveria ser capaz de armazenar dados dos pacientes, nome, idade, gênero, etc, além de dados clínicos como sintomas. Em tal abordagem o modelo conceitual tem relaciona cada paciente a uma amostra. O foco do sistema deixa de ser simplesmente o uso de dados para o treinamento do algoritmo de machine learning, sendo que este último se torna parte de um conjunto de requisitos para implementar o diagnóstico de câncer em pacientes, por exemplo de um hospital.

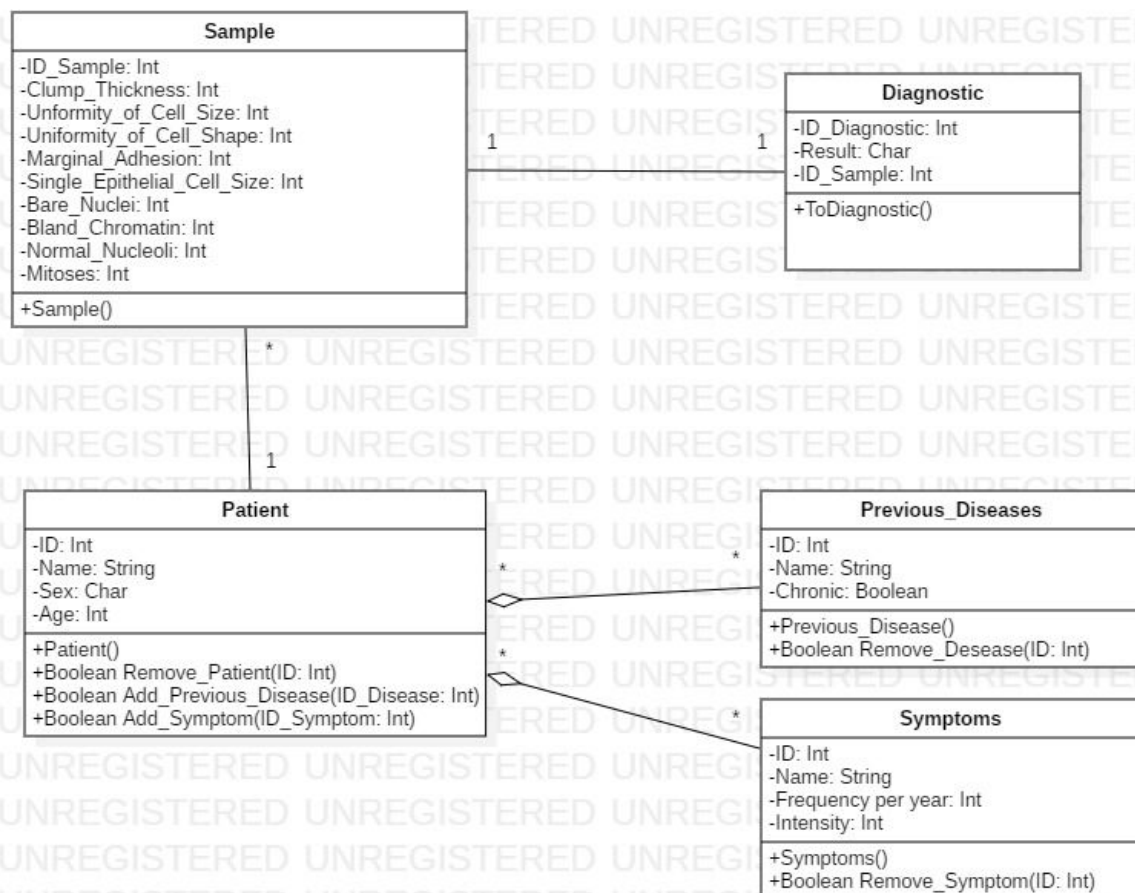


figura 2: modelagem conceitual para a segunda parte do trabalho.

Modelo Lógico:

Na implementação do modelo lógico, cada classe foi modelada como uma tabela, além disso, para representar as relações de “muitos para muitos” foi necessário criar tabelas associativas como por exemplo Patient-Previous_Diseases, que conecta o paciente a seu histórico de doenças.

Uma outra diferença do modelo lógico para o modelo conceitual foi que, apesar de na abordagem conceitual a class Patient se relacionar com cardinalidade (1 -N) com Sample. No modelo lógico, como as amostras da base de dado escolhida não continham um campo para Id do paciente, optou-se por implementar uma relação de cardinalidade (1-1), na qual a tabela paciente contém uma chave estrangeira para a sample (amostra) correspondente a seu exame.

Sample(_ID_Sample_, Clump_Thickness, Uniformity_of_Cell_Size, Uniformity_of_Cell_Shape, Marginal_Adhesion, Single_Epithelial_Cell_Size, Bare_Nuclei, Bland_Chromatin, Normal_Nucleoli, Mitoses, ID_Dianostic, ID_Patient);

CHE: ID_Dianostic (Diagnostic: ID_Diagnostic)

CHE: ID_Patient (Patient: ID);

Diagnostic (_ID_Diagnostic_, Result);

Patient (_ID_, Name, Sex, Age);

Patient-Previous_Diseases(_ID_Patient_, _ID_Previous_Diseases_);

CHE: ID_Previous_Diseases (Previous_Diseases: ID)

CHE: ID_Patient (Patient: ID);

Patient-Symptoms (_ID_Patient_, IDSymptoms_);

CHE: ID_Symptoms (Symptoms: ID)

CHE: ID_Patient (Patient: ID);

Previous_Diseases (_ID_, Name, Chronic);

Symptoms(ID, Name, Frequency_per_year, Intensity);

Código:

A implementação em SQL pode ser encontrada em

<https://github.com/ntak1/404NotFound>