



Instituto Superior de Engenharia de Lisboa

Aprendizagem e Mineração de Dados

Mestrado Engenharia Informática e Multimédia

Trabalho Final

Semestre de Inverno, 2022/2023

Grupo 2

Nome: Gabriel Soares || **Número:** A50231

Nome: Gonçalo Fonseca || **Número:** A50185

Nome: Ruben Albuquerque || **Número:** 49063

Índice

1	Exercicio 1 - Descrição do Cenário	1
2	Exercicio 3 - Análise do Dataset	1
3	Exercicio 4 - Implementação de bases de dados	2
4	Exercicio 5 - Exportação de dados	3

1 Exercício 1 - Descrição do Cenário

O centro médico "**MedKnow**" utiliza um sistema de gestão de bases de dados (**SGBD**) que contém todos os dados recolhidos, ao longo do tempo, sobre a visita de cada paciente a um médico (a trabalhar no "**MedKnow**"). O objectivo atual da equipa de oftalmologia é analisar toda a informação acumulada (ao longo do tempo) para extrair os padrões que fornecem indicadores úteis para apoiar a atividade de prescrição (e diagnóstico).

Para atingir este objectivo, decidiram contactar a empresa "**SoftKnow**" e enviar-lhes um ficheiro contendo um conjunto de dados (relacionado com a actividade de prescrição de lentes) e propuseram o seguinte desafio: "(...)Enviar-nos um protótipo de um sistema que fornecesse à "**MedKnow**" não só o apoio operacional (trabalho diário), mas também a perspectiva estratégica (padrões úteis) que podem ser extraídos a partir destes dados de trabalho diário". Neste primeiro cenário, vamos experimentar um conjunto de dados muito simples e pequeno.

2 Exercício 3 - Análise do Dataset

O conjunto de dados enviado pela "**MedKnow**" contém todos os dados recolhidos, ao longo do tempo, sobre a visita de cada paciente a um médico da sua equipa de oftalmologia. Contém quatro condições específicas do doente (atributos variáveis) e o tipo de lentes que foram prescritas a esse doente (o atributo da etiqueta de classe).

O conjunto de dados contém 5 atributos e 16 registos diferentes.

age - A idade do paciente (discreta). Pode ser atribuída a 3 valores:

1. "young- quando o paciente ainda é jovem (normalmente 0 - 27 anos);
2. "pre-presbyopic- quando o paciente já não é jovem, mas ainda não desenvolveu qualquer condição presbiopia ou vista cansada (normalmente 28 - 38);
3. "prebyopic- quando o paciente desenvolveu uma condição presbitópica. Note-se que a presbiopia é uma condição ocular em que o olho do paciente perde lentamente a capacidade de se concentrar rapidamente em objectos que estão próximos. É uma doença que afecta toda a gente durante o processo natural de envelhecimento.

prescription - O tipo de defeito ocular do paciente (discreto). Pode ser atribuído 2 valores:

1. "hypermetrope- quando o doente pode ver objectos distantes mas não consegue ver claramente os objectos próximos (hipermetropia);
2. "myope- quando o doente é incapaz de ver as coisas claramente, a menos que estejam relativamente perto dos olhos (miopia).

astigmatic: Se o paciente sofre de astigmatismo ou não (discreto). Podem ser atribuídos 2 valores: "sim" ou "não". Note-se que o astigmatismo é uma imperfeição comum e geralmente tratável na curvatura do olho, que causa distância desfocada e visão próxima.

tear_rate: A taxa de laceração do paciente (discreta). Pode ser atribuída a 2 valores: "normal" ou "reduced" (reduzido). Note-se que a taxa de laceração é definida como a diminuição percentual por minuto da concentração de fluoresceína nas lacerações do paciente após a instilação da fluoresceína.

lenses: O tipo de lentes prescritas ao doente, com base nos outros atributos (etiqueta de classe, discreta). Podem ser atribuídos 3 valores:

1. "hard" se o paciente for mais aconselhado a usar lentes de contacto duras;
2. "soft" se o paciente for mais aconselhado a usar lentes de contacto macias;
3. "none" se o doente não precisar de lentes.

3 Exercício 4 - Implementação de bases de dados

Em primeiro lugar, uma base de dados relacional precisa de ser concebida para apoiar o trabalho diário no hospital. Como é ilustrado pelo diagrama **Entidade-Relação** na **figura 1**, as entidades envolvidas consistirão em **Doctor**, **Disease**, **Patient** e **Diagnostic**.

Patient: O indivíduo a ser examinado pelo Doutor (**Doctor**). Tem um nome, um **número de CC** (Cartão de Cidadão) e uma **data de nascimento**.

Doctor: O indivíduo que irá tratar o Paciente (**Patient**). Tem também um nome, um **número de CC** (Cartão de Cidadão) e uma **data de nascimento**.

Diagnostic: Após o exame, o Doutor (**Doctor**) chega a algumas conclusões sobre o estado de saúde visual do Paciente. Esta entidade será constituída por uma tabela com todas as combinações possíveis dos valores que cada atributo (**idade** e **taxa de laceração**) pode assumir, cada combinação identificada por um id.

Disease: Com o conhecimento sobre a condição do Doente - (Diagnóstico - **Diagnostic**) -, o Doutor (**Doctor**) identificará em que situação da doença se enquadra. Tal como a entidade Diagnóstica (**Diagnostic**), esta também será constituída por uma tabela contendo todas as combinações possíveis de cada doença (**isAstigmatic**, **isMyope** e **isHypermetrope**), cada combinação identificada pelo seu **id**.

Como é possível que o mesmo **Patient** possa ser **astigmático**, **miope** e/ou **hipermetrope**, as três condições foram separadas e são agora independentes. Cada uma delas pode assumir "verdadeiro" ou "falso" como valores possíveis.

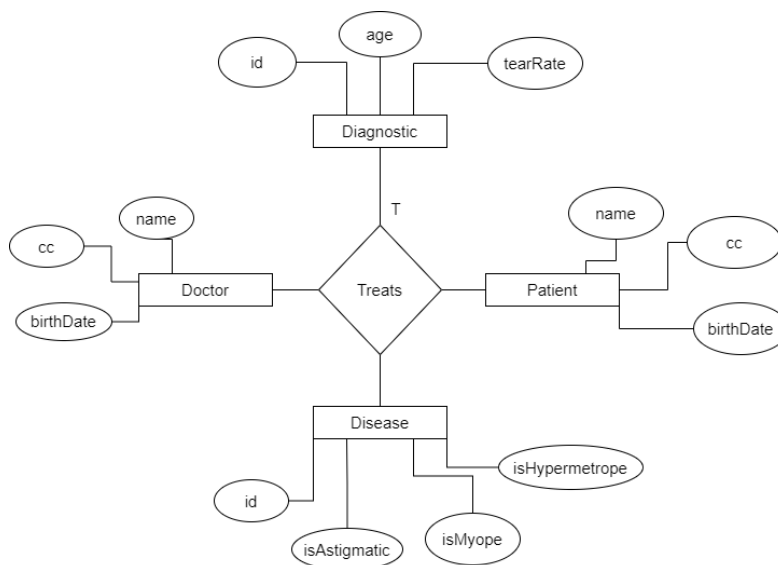


Figura 1: Diagrama Entidade-Relação

Note-se que existe uma relação quaternária chamada **Treats** (Tratamentos) que irá relacionar as quatro entidades anteriormente descritas. Esta relação será a nossa quinta tabela.

Agora, o esquema da base de dados precisa de ser implementado utilizando SQL. A pasta 'scripts' contém os scripts utilizados para criar e povoar a base de dados.

4 Exercício 5 - Exportação de dados

Para exportar uma selecção de dados a serem analisados, usaremos o comando **SQL** de **CREATE VIEW** para agrupar os dados de que precisamos, e o comando **SQL** de **COPY** para os exportar para um ficheiro externo.

Algo que temos de considerar é o facto de que os dados precisam de ser exportados num formato adequado para a framework **Orange**.

Cada tabela de uma base de dados relacional contém pelo menos um cabeçalho, com o nome dos atributos na tabela. A estrutura Laranja precisa de dois cabeçalhos adicionais: Um segundo cabeçalho indicando se o atributo assume valores discretos ou contínuos, e um terceiro cabeçalho indicando que atributo é a etiqueta de classe.

A selecção de dados que iremos exportar contém apenas informações úteis para

a prescrição das lentes. Por exemplo, o **número CC** e o nome do **Patient** (Paciente) não são importantes neste contexto. Aqui estão os cabeçalhos da vista criada:

age	tearRate	isAstigmatic	isMyope	isHypermetrope	prescribedLenses
discrete	discrete	discrete	discrete	discrete	discrete
					class

Agora, com o comando **COPY**, esta **view** será exportada para um ficheiro **txt** externo, contendo os cabeçalhos previamente descritos e todos os registos.