1. **Análise do excerto de dados e suposições sobre a atividade diária da “MedKnow”**

O ficheiro disponibilizado pela MedKnow apresenta um conjunto de dados relacionados com a prescrição de lentes oftalmológicas, onde cada registo (instância) representa uma consulta de um paciente. As variáveis disponiveis no conjunto de dados descrevem as principais caracteristicas observadas e recolhidas durante o exame oftalmológico, nomeadamente: a idade dos olhos, o tipo de visão (que seria a doença), a presença ou ausência de astigmatismo, a frequência de lágrimas e, por ultimo, o tipo de lente prescrita pelo médico com base nas variáveis anteriores.

Com base nos dados deste ficheiro assume-se que a atividade diária desta empresa seja composta pelos seguintes pontos:

* Realização de consultas oftalmológicas que servem para recolher dados sobre cada paciente;
* Registo de todos estes dados numa base de dados, permitindo o acompanhamento histórico de cada paciente.
* A análise dos dados recolhidos durante as consultas de maneira a prescrever um tipo de lentes para cada doente.

Posto isto, e com o objetivo de automatizar este ultimo ponto, o objetivo deste projeto é desenvolver um protótipo de um sistema capaz de transformar dados de consultas em conhecimento útil, permitindo predizer um tipo de lentes para um paciente de forma mais rápida e automática. Para este fim, o sistema irá recorrer a técnicas de aprendizagem automática e mineração de dados, treinando com dados existentes para aprender padrões de decisão e, assim, ser capaz de predizer o tipo de lentes com base nas informações recolhidas nas consultas.

1. **Modelo de dados (EA)**

Com base na informação que a empresa utiliza uma base de dados relacional que suporta, pelo menos, as entidades *PATIENT*, *DOCTOR*,e *DISEASE*, e os atributos *birthDate* e *diseaseName*, tendo este ultimo três valores possiveis presentes no ficheiro disponibilizado - sendo estes: *myope*, *hypermetrope* e *astigmatic* - e de maneira a suportar as suposições sobre a atividade diária da empresa, desenvolveu-se o seguinte modelo de dados, neste caso Entidade-Associação.

A diagram of a company

AI-generated content may be incorrect.

Figura 1 - Modelo EA desenvolvido

1. **Criação e população da base de dados**

A screenshot of a computer program

AI-generated content may be incorrect.

Figura 2 - Ficheiro sql para criação das tabelas

A população das tabelas foi feita com dados predefinidos aleatórios para teste. Apenas a tabela das consultas foi populada em conformidade com o ficheiro disponibilizado pela empresa, ou seja, com as mesmas instâncias associadas a um paciente e doutor aleatórios, tendo ainda uma data de consulta associada.

A screen shot of a computer screen

AI-generated content may be incorrect.

Figura 3 - População da tabela de consultas

1. **Exportação de dados**

A exportação de dados foi feita de maneira a ter um formato compativel com o framework do orange, sendo este formato o 3RowHeader.

Foram criadas três *views*, sendo a primeira para identificar a class, denominada lenses\_class, a segunda para determinar o tipo de dados, neste caso todos discretos, denominada lenses\_domain e a terceira para recolher os dados relevantes da consulta, como as caracteristicas do paciente e as lentes escolhidas.

Após a criação dessas três views, foram unidas em apenas uma view denominada lenses\_dataset que estará no formato 3RowHeader pronta para ser exportada.

A screenshot of a computer program

AI-generated content may be incorrect.

Figura 4 – Criação das views que compoêm o 3RowHeader

A computer screen with white text

AI-generated content may be incorrect.

Figura 5 - Criação da view com o formato 3RowHeader que será exportada

A computer screen with text on it

AI-generated content may be incorrect.

Figura 6 - Exportação da view

A screen shot of a computer

AI-generated content may be incorrect.

Figura 7 - Lenses\_dataset view