Instituto Superior de Engenharia de Lisboa

**Final Project A**

**Aprendizagem e Mineração de Dados**

Mestrado em Engenharia Informática e de Computadores

Docente: Paulo Trigo

Grupo 25

Nuno Gomes – 44021

Índice

[Introdução 3](#_Toc59061908)

[Análise de dados 4](#_Toc59061909)

[Modelo EA 5](#_Toc59061910)

[Modelo Relacional 6](#_Toc59061911)

[Legenda 6](#_Toc59061912)

[Tabelas 6](#_Toc59061913)

[Regras de Negócio: 6](#_Toc59061914)

[Procedimento Experimental 7](#_Toc59061915)

[Resultados 8](#_Toc59061916)

[Análise de Resultados 9](#_Toc59061917)

[Conclusão 10](#_Toc59061918)

# Introdução

Métodos de classificação – intro a partir daqui

Falar do 1R

Falar do ID3

Falar do Naive Bayes

Treino e teste dos modelos

Falar dos métodos de treino e teste

Validacao de modelos

Sampling as cross validation, random sampling, leave one out e chega, falar por alto test and train data e falar de overfitting

Falar Dos dados e das métricas – F1, Accuracy, Precision, Recall  
e Matriz de confusão

# Análise de dados

No documento que nos foi providenciado para analisar (d01\_lenses.xls), analisámos os dados para inferir qual o trabalho diário realizado nesta empresa. Com base na nossa análise, compreendemos que o trabalho se foca maioritariamente na área da oftalmologia, e conseguimos tirar as seguintes conclusões mais específicas:

- Medição de indicadores de saúde ocular, de forma a posteriormente auxiliar uma escolha mais informada de um tratamento, no caso de ser necessário.

- Diagnóstico de doenças oculares, especificamente as três doenças oculares mais comuns no geral da população que são hipermetropia, miopia e astigmatismo.

- Elaboração de tratamento. No caso de ser diagnosticada uma doença ocular a um paciente será proposto um tratamento. O tratamento atribuído serão lentes oculares para correção de visão. Estas são identificadas pela sua rigidez, das quais as mais populares são as rígidas e as moles.

Foram feitas algumas assunções em relação aos dados.

* Sempre que há contacto entre um médico da Medknow e um paciente, é gerada uma avaliação do paciente com os dados atualizados das suas condições oftalmológicas.
* Um paciente pode ser diagnosticado com mais que uma doença.
* É assumido que as idades oculares correspondem a: jovem, desde os 0 aos 34 anos; pré-presbiópico dos 35 aos 44 anos; presbiópico com mais de 45 anos.
* Foi assumido que há mais de três tipos de doença, bem como mais de dois tipos de lente
* Não há uma relação de acompanhamento entre médico e doente explicita.

# Modelo EA

# Modelo Relacional

## Legenda

**CE** = Chaves Estrangeiras

**CC** = Chaves Candidatas

**RI** = Regras de Integração

## Tabelas

*Patient* (person\_id, birthdate, patient\_Name)

**RI: birth\_date é escrito na forma 'yyyy-mm-dd'**

*Disease* (disease\_name)

*Diagnosis* (patient\_id, disease\_name, diagnosis\_date)

**CE:** (patient\_id ) referencia *Patient*(person\_id),

(disease\_name referencia *Disease*(disease\_name)

*Doctor* (person\_id, internal\_id, birthdate, doctor\_name)

**CC:** internal\_id

**RI:** internal\_id é um valor único.

*Lenses*(lens\_type)

*Evaluation* (doctor\_id,patient\_id,lens\_type, ocular\_age, tear\_rate, evaluation\_date)

**CE:** (patient\_id ) referencia *Patient*(person\_id),

(doctor\_id ) referencia *Doctor*(doctor\_id),

(lens\_type) referencia *Lenses*(lens\_type)

**RI:** ocular\_age apenas toma por valores [young, pre-presbyopic, presbyopic]

### Regras de Negócio:

O atributo person\_id utilizado nas tabelas *Doctor* e *Person* representa o número de identificação pessoal de cada uma destas pessoas. Na tabela *Doctor*, o internal\_id é o número do médico dentro da empresa “Medknow”.

Na tabela *Lenses*, o atributo lens\_type foi intencionalmente deixado com os valores em aberto, pois após uma pesquisa foi descoberto que há mais tipos de lentes do que aqueles utilizados nos dados que nos foram passados.

Por fim, a tabela *Evaluation*, não costringe o valor de tear\_rate a [normal, reduced], pois o grupo também acha que há mais valores de tear\_rate que possam ser admitidos.

# Procedimento Experimental

Foca-se maioritariamente nos exercícios 6, 7 e 8

# Resultados

# Análise de Resultados

Aqui faz-se o exercício 9 - Provide “MedKnow” an evaluation of both methods and your own conclusions on the most reliable approach – 1R vs Bayes vs ID3

# Conclusão