**Objectivos**

* Descrever a realização de experimentos de Aprendizagem Automática no Weka.
* Descrever os conjuntos de dados dos arquivos *weather.nominal.arff*, *iris.arff*, *contact-lens.arff*, *german-credit.arff*, *glass.arff* e *ionosphere.arff*.
* Descrever algumas operações de preprocessamento aplicados sobre os dados.
* Explicar dois métodos distintos, ID3 e C4.5, utilizados para a aprendizagem de árvores de decisão.

**Introdução**

Este informe irá apresentar algumas técnicas de Aprendizagem Automática experimentadas nas aulas de AA: operações de preprocessamento dos dados e a aprendizagem de árvores de decisão usando como ferramenta de suporte o Weka.

Os arquivos *weather.nominal.arff*, *iris.arff*, *contact-lens.arff*, *german-credit.arff*, *glass.arff* e *ionosphere.arff* são usados como os conjuntos de dados para a aprendizagem de conceitos.

Temos por pré-requisitos para a aprendizagem, o estudo de diversos aspectos:

* Operações de pré-processamento
  + Normalização
  + Discretização
* Métodos de validação
  + Medidas de desempenho,
  + Holdout
  + Validação cruzada
  + Matriz de confusão
* Árvores de decisão

Nas próximas páginas iremos fazer uma breve abordagem de cada aspecto.

**Conjunto De Dados**

***weather.nominal.arff:***

**Conceito retratado:** Possibilidade para jogar (sim/não)

**Quantidade de instâncias:** 24

**Quantidade de atributos:** 5

**Quantidade de classes:** 2

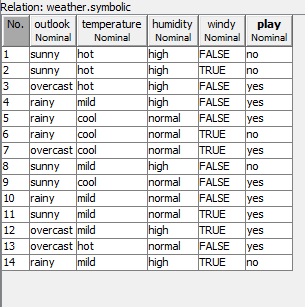
******

Figura Conjunto de Dados weather.nominal

***contact-lens.arff*:**

**Conceito retratado:** Possibilidade de o paciente se ajustar a uma lente de contacto

**Quantidade de instâncias:** 24

**Quantidade de atributos:** 4

**Quantidade de classes:** 3

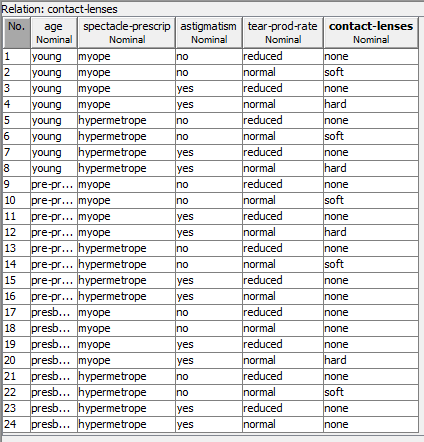


Figura Conjunto de Dados contact-lenses

***german-credit.arff* :**

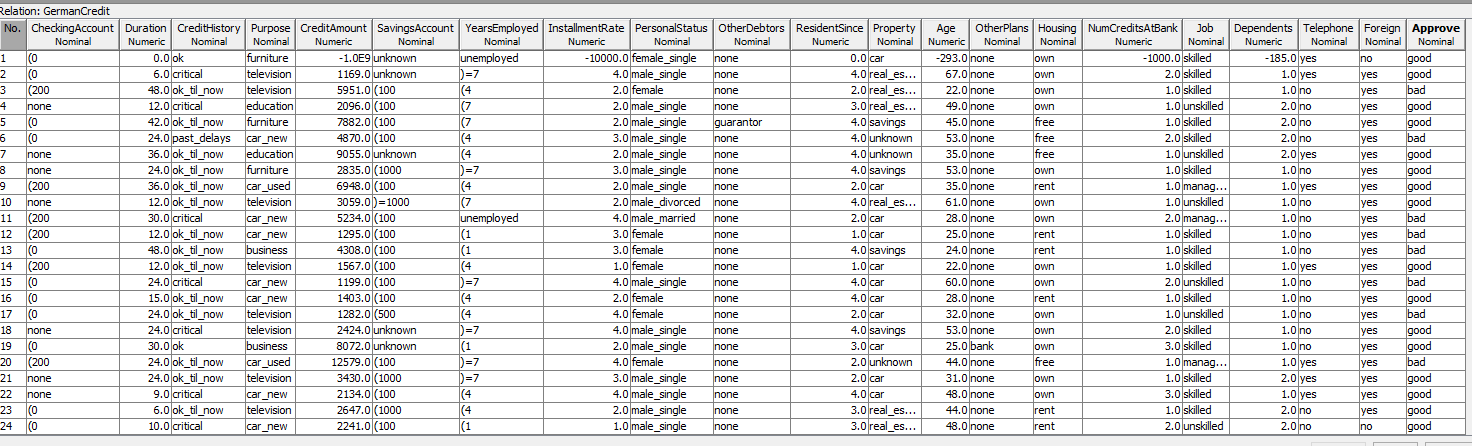
**[](credit.PNG)**

Figura Conjunto de dados german-credit com excerto das instâncias de 1-24; O total 1001 instâncias

***Obs: link inserido para ver a figura com melhor resolução; ctrl+click na foto.***

**glass.arff:**

**Conceito retratado:** Identificação de vidro

**Quantidade de instâncias:** 214

**Quantidade de atributos:** 10

**Quantidade de classes:** 7

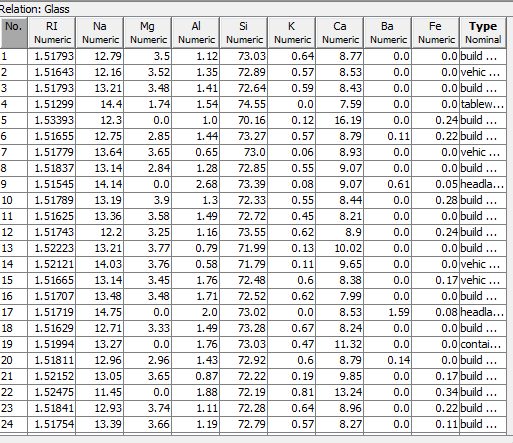


Figura Conjunto de dados glass com excerto das instâncias de 1-24; O total 214 instâncias

***ionosphere.arff:***

**Conceito retratado:** Identificação de vidro

**Quantidade de instâncias:** 351

**Quantidade de atributos:** 35

**Quantidade de classes: 2**

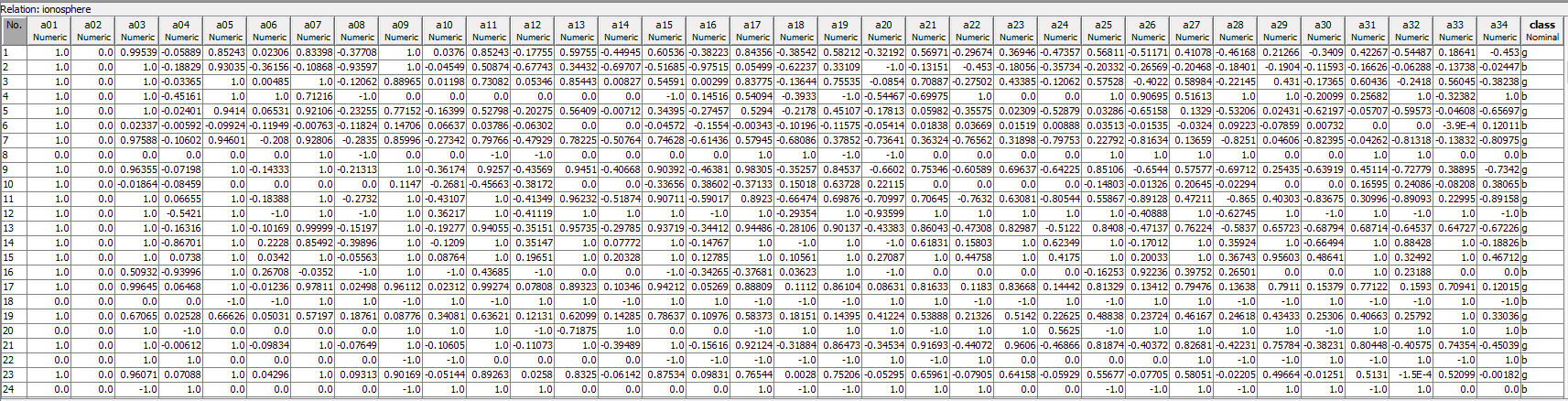
**[](ionosphere.PNG)**

Figura Conjunto de dados ionosphere com excerto das instâncias de 1-24; O total 351 instâncias

***Obs: link inserido para ver a figura com melhor resolução; ctrl+click na foto.***

**Pré-Processamento**

Em muitos casos é necessário efectuar certas operações de processamento aos dados antes da sua utilização,***Pré-Processamento.***

Entre essas operações temos:

– A normalização dos dados

– Discretização dos dados

**Normalização dos dados**

* Conceitos teóricos importantes relacionados
* Breve descrição dos conjuntos de dados utilizados (conceito retratado, quantidade de instâncias, quantidade de atributos, quantidade de classes...)