

***Elementos RUP***

•

Papeis: define o comportamento e as responsabilidades de um determinadoindivíduo ou grupo de indivíduos trabalhando como uma equipe. Papéis não sãoindivíduos e nem títulos de trabalho. Um indivíduo pode assumir vários papéis. Sãoexemplos de papéis:

*o*

Analista de sistema - O indivíduo que assume este papel coordena aobtenção dos requisitos e a modelagem dos casos de uso identificandofuncionalidades do sistema e estabelecendo limites do sistema;

*o*

Projetista - Esse indivíduo define responsabilidades, operações, atributos,relacionamentos de uma ou mais classes e determina como elas devem serajustadas para serem implementadas no ambiente;

*o*

Projetista de testes - Responsável pelo planejamento, projeto, implantação eavaliação de testes, incluindo a geração de plano e modelo de teste,

implementando procedimentos de testes e avaliando a abrangência dostestes, resultados e a efetividade.

•

Uma atividade é uma unidade de trabalho que um individuo executa quando estáexercendo um determinado papel e produz um resultado importante para o contextodo projeto. Cada atividade pode ser dividida em passos. São exemplos deatividades: planejar uma iteração, encontrar casos de uso e atores, rever o projeto,executar um teste de performance.

•

Um artefato é um pedaço de informação que é produzido, modificado ou utilizadoem um processo. Os artefatos são os produtos de um projeto, são utilizados comoentradas de atividades e são produzidos como saída. São exemplos:

*o*

Um modelo, como um modelo de caso de uso

*o*

Um elemento de um modelo, como uma classe

*o*

Um documento

*o*

Um código fonte

*o*

Executáveis

•

Fluxos de trabalhos são seqüências de atividades que são executadas para aprodução de um resultado valioso para o projeto. Podem ser representados pordiagramas de seqüências, colaboração e atividades da linguagem UML

•

Disciplina é um conjunto de atividades relacionadas que fazem parte de umcontexto comum em um projeto. Proporcionam um melhor entendimento do projetosob o ponto de vista tradicional do modelo cascata. O RUP possui nove disciplinasdivididas em disciplinas de processo (modelagem de negócios, requisitos, análise edesign, implementação, teste e distribuição) e suporte (configuração egerenciamento de mudanças, gerenciamento de projeto e ambiente).

**Universidade Federal de Viçosa**

**Campus Rio Paranaíba**

**Aldo Henrique Dias Mendes**

**Matrícula 1610**

**Rafael Martins**

**Matrícula 1566**

**Michel Junio Ferreira Rosa**

**Matrícula 1540**

**Mineração de Dados**

**Classificação de Dados**

**Rio Paranaíba, Minas Gerais**

**06/2014**

Sumário

[1. Base de dados estação UFV-CRP 3](#_Toc391223936)

[1.1 Resultados 3](#_Toc391223937)

[2. Base de dados estação INMET Patrocínio 5](#_Toc391223938)

[2.1 Resultados 6](#_Toc391223939)

[3. Conjunto de dados de lentes 13](#_Toc391223940)

[3.1 Pre-Processamento 13](#_Toc391223941)

[3.2 Resultados 14](#_Toc391223942)

# Base de dados estação UFV-CRP

Base de dados gerada pela estação climática localizada no campus de Rio Paranaíba da Universidade Federal de Viçosa, está capturando dados desde 13/06/11 e a captura dos dados acontece de hora em hora. Até a data 04/06/2014 são aproximadamente 25000 objetos e 62 atributos + 1 classe que foi criada.

Entre os 62 atributos foram escolhidos alguns em específicos que são: Umidade, Velocidade do vento, Chuva, Radiação, Evapotranspiração. A escolha ocorreu através da referência bibliográfica e baseado em gráficos que mostraram que estes atributos são relevantes para determina o atributo escolhido como classe que foi a Temperatura.

Os objetos que possuem algum atributo faltante foram simplesmente retirado da base, uma vez que ao substituir o dado faltante por 0, significaria uma grande perda nos resultados finais.

Como este atributo está na forma de valores contínuos foi necessário transformá-lo em um atributo nominal, dessa forma foi utilizado um algoritmo para conversão representado da seguinte forma:

**if(temperatura<18)**

**classe="Frio";**

**if(temperatura>=18 && temperatura<25)**

**classe="Normal";**

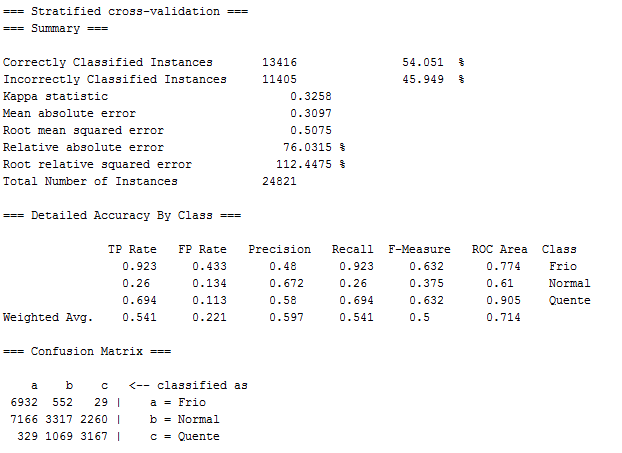
**if(temperatura>=25)**

**classe="Quente";**

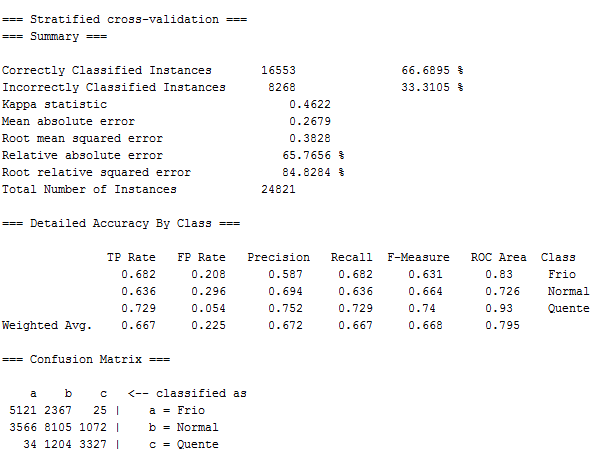
Para classificação foi utilizado o *software* WEKA, e os algoritmos J48 que é a implementação da arvore de decisão e o algoritmo NaiveBayesSimple que é a implementação do Bayesiano Simples.

Do dia 04/06/2014 até o dia 15/06/2014, foram capturado dados para serem classificados nos algoritmos treinados com os dados anteriores, entre esses dados não possuem dados faltantes, o que ajuda na sua utilização.

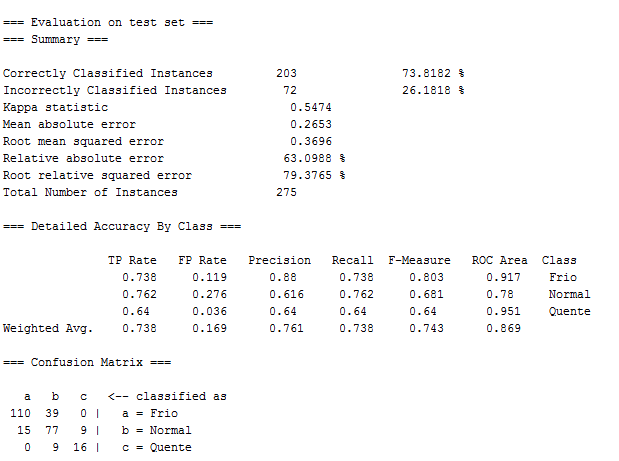
## Resultados



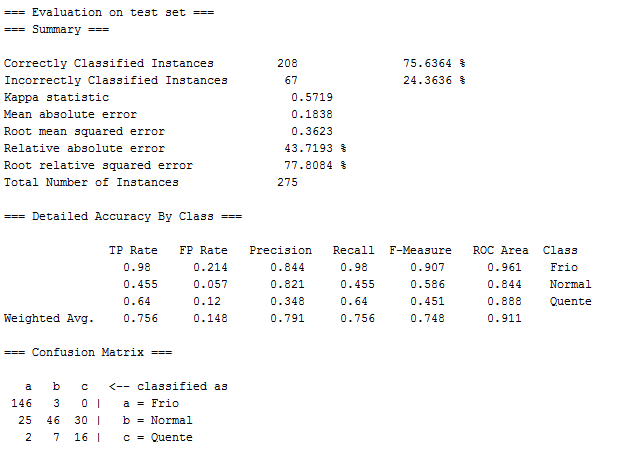
Bayesiano NaiveBayesSimple utilizando **Cross-validation = 10.**



Árvore J48 utilizando **Cross-validation = 10.**



Árvore J48 utilizando **Novos objetos.**



Bayesiano NaiveBayesSimple utilizando **Novos objetos**.

Árvore J48

Bayesiano NaiveBayesSimple

# Base de dados estação INMET Patrocínio

Base de dados gerada pela estação climática PATROCINIO-A523 localizada na cidade de Patrocínio e fornecida pelo Instituto Nacional de Meteorologia - INMET. A coleta dos dados é feita minuto a minuto e integralizada na base de dados a cada hora. É disponível para download somente os dados dos últimos 90 dias, desta forma, o intervalo dos dados coletados foi de 17/03/2014 até 15/06/2014, totalizando 2181 objetos. A base é composta por 17 atributos + 1 classe que foi criada.

Entre os 17 atributos, foram escolhidos alguns em específicos. São eles: Umidade, Pressão, Direção do vendo, Velocidade do vento e Radiação. A escolha ocorreu através de estudos bibliográficos e gráficos referentes aos atributos escolhidos, que mostram ser relevantes ao determinar o atributo escolhido como classe que foi a Temperatura.

O atributo Temperatura era representado através de dados contínuos, desta forma foi necessário transformá-lo em um atributo nominal, através da utilização de um algoritmo para conversão representado da seguinte forma:

**If (temperatura < 18)**

**classe = "Frio";**

**if (temperatura >= 18 && temperatura < 25)**

**classe = "Normal";**

**if (temperatura >= 25)**

**classe = "Quente";**

Vale ressaltar, que todos os objetos foram normalizados para o intervalo de [0 a 1].

Para classificação foi utilizado o *software* WEKA, o algoritmo J48, que é a implementação da árvore de decisão e o algoritmo NaiveBayesSimple que é a implementação do Classificador Bayesiano Simples.

Após o treinamento dos algoritmos com a base de dados: Patrocínio, foram realizados testes de classificação com uma base de dados da cidade de Araxá, sendo que a mesma, possui o mesmo número de objetos e foi pré-processada com as mesmas especificações da base de treinamento.

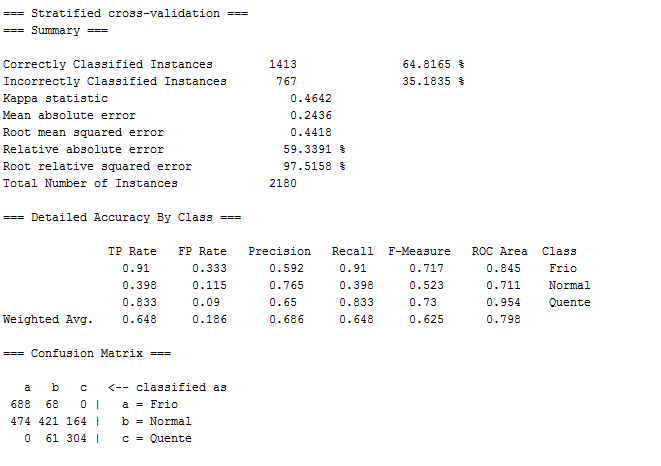
Os resultados podem ser avaliados na seção abaixo.

## 2.1 Resultados

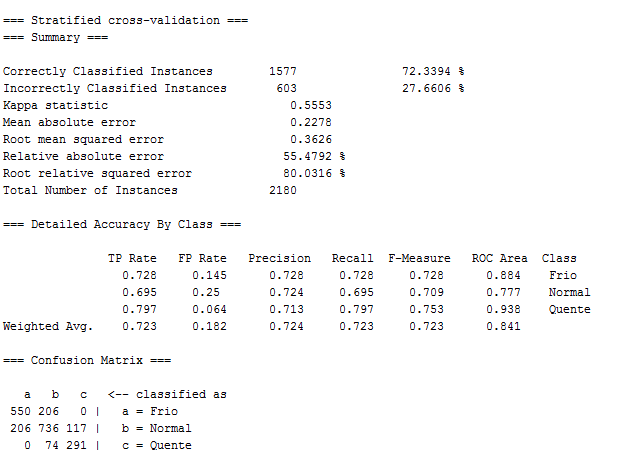
Gráficos gerados através dos objetos da base, para auxílio da interpretação dos dados

**Análises obtidas a partir do WEKA**

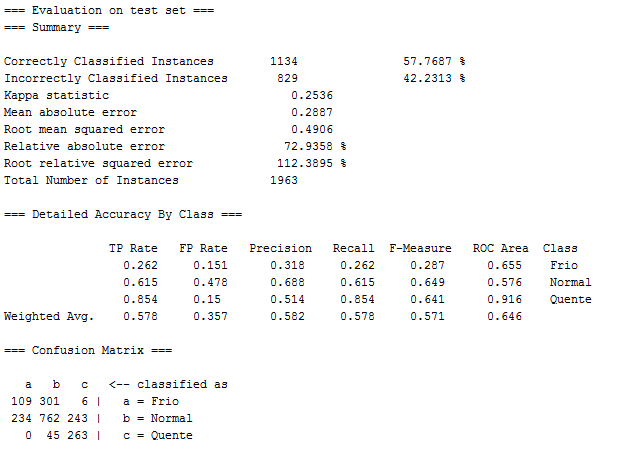
Bayesiano NaiveBayesSimple utilizando **Cross-validation = 10.**



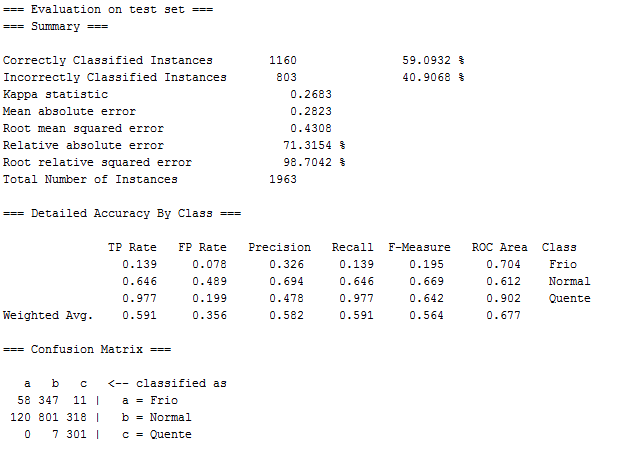
Árvore J48 utilizando **Cross-validation = 10.**



Bayesiano NaiveBayesSimple utilizando **Novos objetos**.



Árvore J48 utilizando **Novos objetos.**

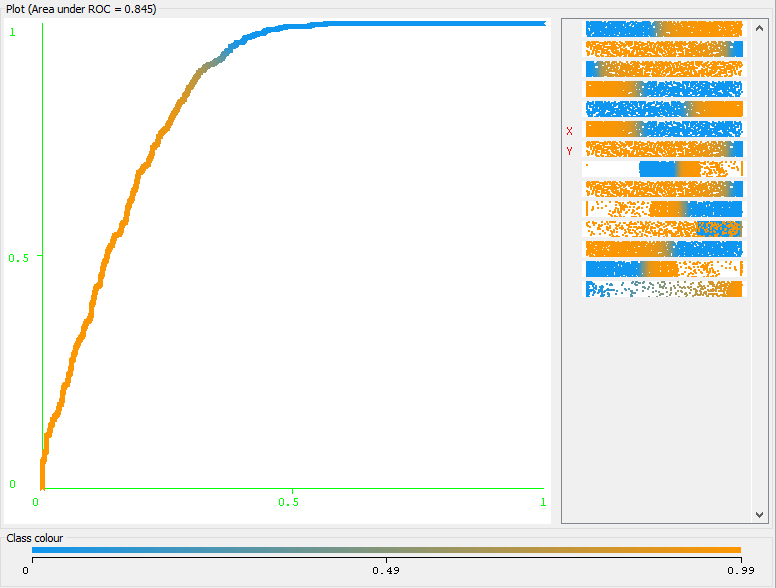


Árvore J48

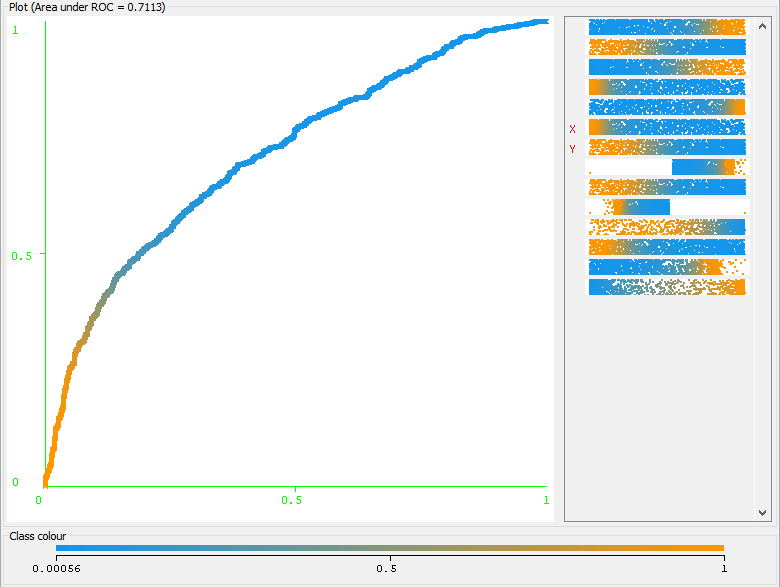
Bayesiano NaiveBayesSimple

**Curvas ROC - Bayesiano NaiveBayesSimple**

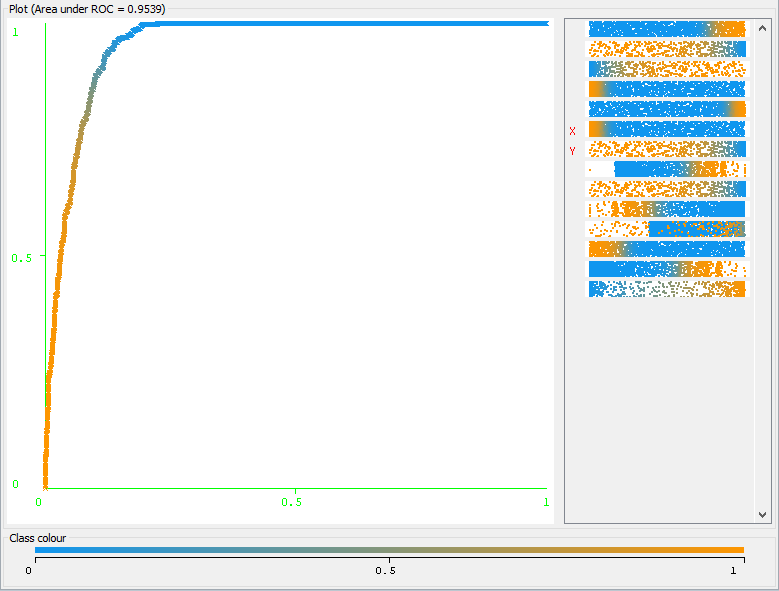
**Classe: Frio**

****

**Classe: Normal**

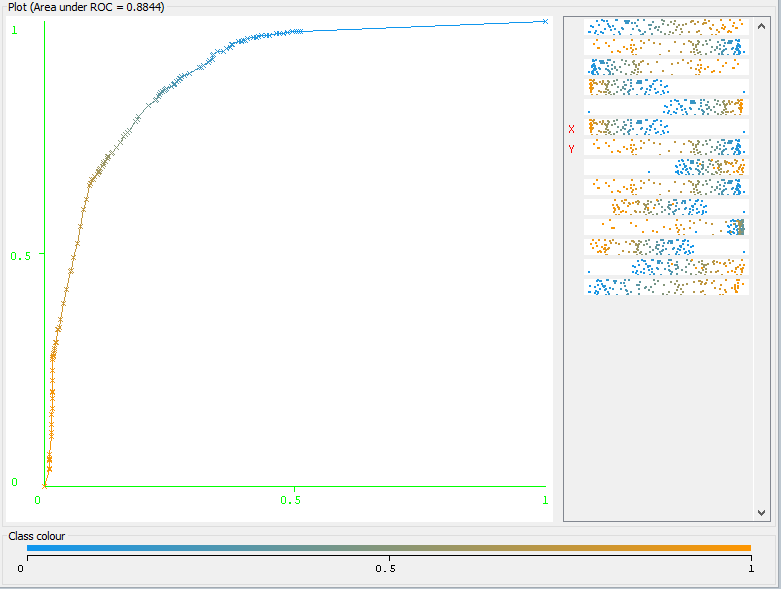
****

**Classe: Quente**

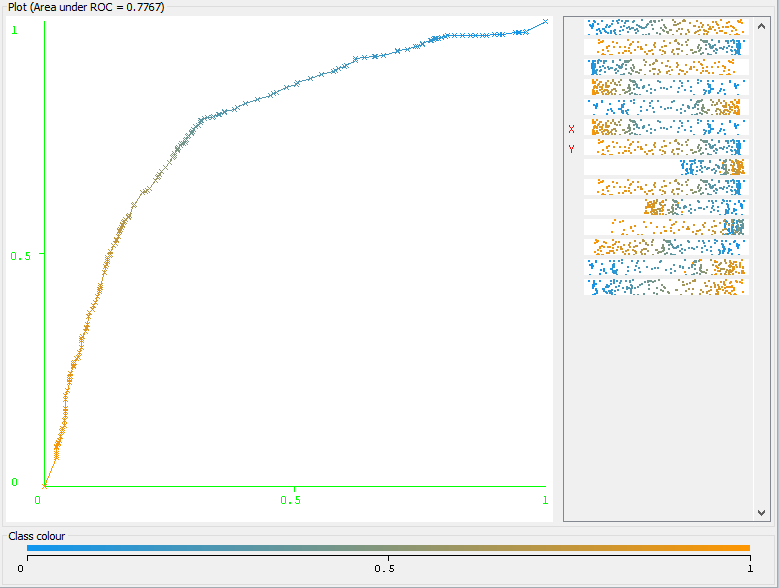
****

**Curvas ROC - Árvore J48**

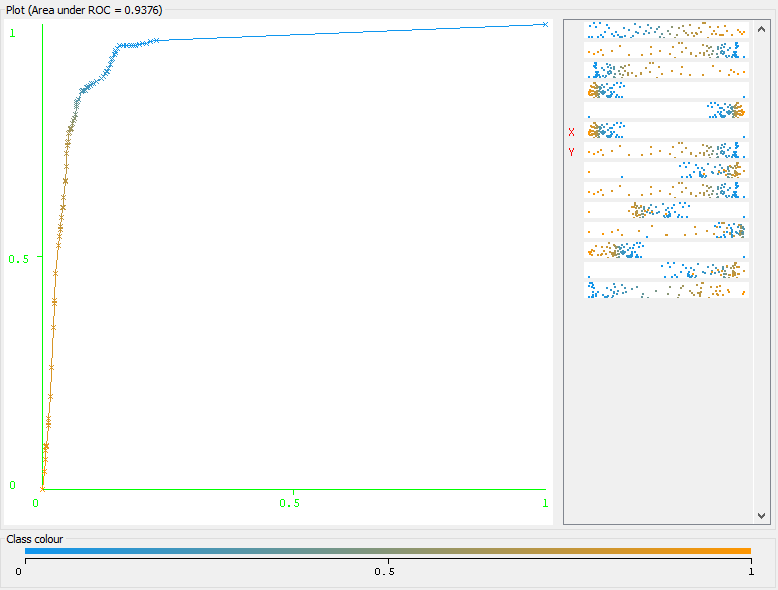
**Classe: Frio**

****

**Classe: Normal**

****

**Classe: Quente**

****

É possível observar com a base de dados testada, que a partir da classe Quente, o classificador bayesiano obteve um melhor desempenho em relação a árvore de decisão, porém as outras duas classes: Frio e Normal, indica uma melhor performance ao executar sobre a árvore de decisão.

# Conjunto de dados de lentes

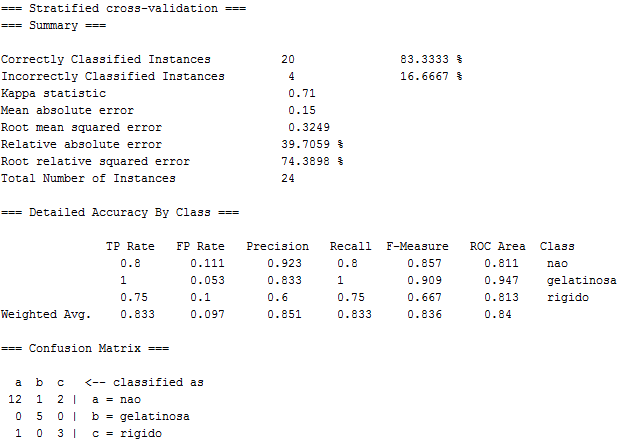
Base de dados que tem como finalidade verificar se o cliente pode ou não usar lentes de contato, se sim qual lente de contato especifica o cliente pode usar, exemplo lentes rígidas ou gelatinosas. Cada objeto é completa e correta. Sendo que a base é livre de ruídos. E se encontra para download em <http://archive.ics.uci.edu/ml/datasets/Lenses>. A base de dados tem exatamente 24 objetos, foi escolhida pelo número inferior, assim testamos o classificador bayesiano se é bom também para classificar poucos objetos.

A base possui 4 atributos mais 1 classe no qual o objeto será classificado. O primeiro atributo chamado de idade das vistas tem como classificação três opções: novo, Pre-presbiopia e Presbiopia. Onde novo é uma visão nova, Presbiopia é uma visão velha e a Pre-presbiopia é o meio termo. O Segundo atributo, a prescrição são problemas que o cliente já possui, como por exemplo: Mópe e Hipermetrope. O terceiro atributo, é o astigmatismo, se a pessoa tem ou não essa falha na visão. E por último o atributo lagrima, se o paciente tem a lagrima reduzida ou é normal.

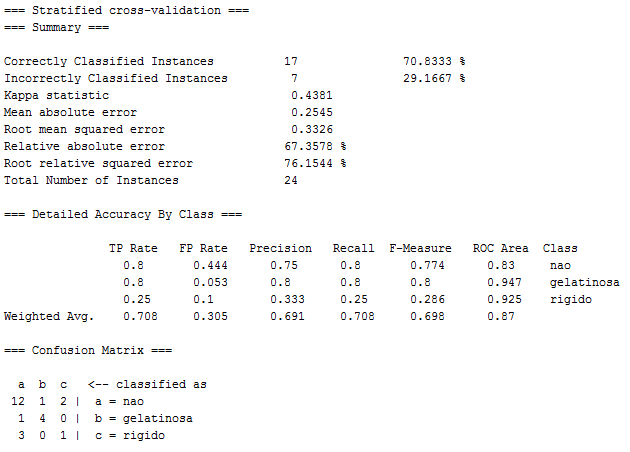
## Pre-Processamento

O único pre-processamento feito a trasformação dos dados da base que estavam com números discretos para atributos nomais, seguindo a documentação da base de dados. Como os atributos nominais estavam completos, ou seja, objetos faltantes e não há a necessidade de normalizar os dados. Foi jogado a base no Weka e abaixo segue os resultados obtidos.

## Resultados

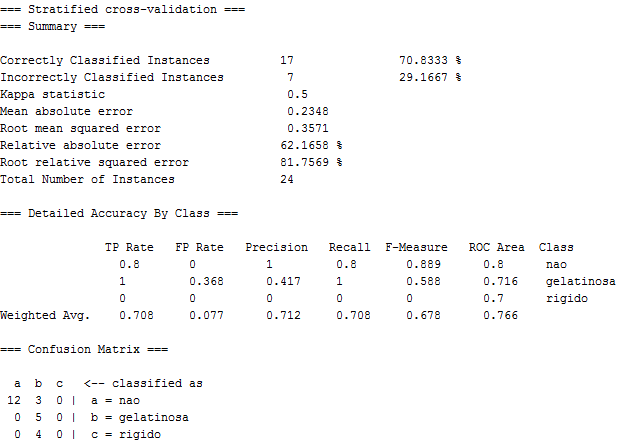


Árvore J48 utilizando **Cross-validation = 10.**

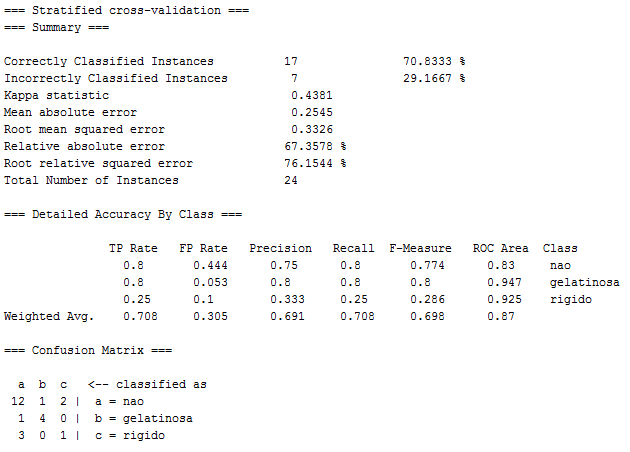


Bayesiano NaiveBayesSimple utilizando **Cross-validation = 10.**

Utilizando todos os atributos o bayesiano obteve uma taxa de acerto de 71% abaixo da árvore que obteve 83%.



Árvore J48 utilizando **Novos objetos.**



Bayesiano NaiveBayesSimple utilizando **Novos objetos**.

Árvore J48

Bayesiano NaiveBayesSimple

Quando pedimos para o WEKA selecionar os atributos e classificar os dois classificados chegaram a uma taxa de acerto de 71%, só que a árvore de decisão passou a não conseguir classificar mais as lentes rígidas, assim perdendo muito o seu desempenho.