|  |  |
| --- | --- |
|  | **FIPP - FACULDADE DE INFORMÁTICA DE PRESIDENTE PRUDENTE**  **Sistema de Informações** |

**Gustavo Akira Felizardo Ito**

**Humberto Caetano Stuani**

**Samuel Felipe Dornelles**

**Manipulação de Dados em R**

Presidente Prudente – SP 2022

**SUMÁRIO**

|  |  |  |
| --- | --- | --- |
| **1** | **INTRODUÇÃO........................................................................................** | **3** |
| 1.1 | Contextualização do problema................................................................ | 3 |
| 1.2 | Objetivo .................................................................................................. | 3 |
| **2** | **ÁRVORE DE DECISÃO..........................................................................** | **3** |
| 2.1 | Procedimentos......................................................................................... | 3 |
| 2.2 | Matriz de confusão................................................................................. | 6 |
| **3** | **APLICAÇÃO............................................................................................** | **7** |
|  |  |  |

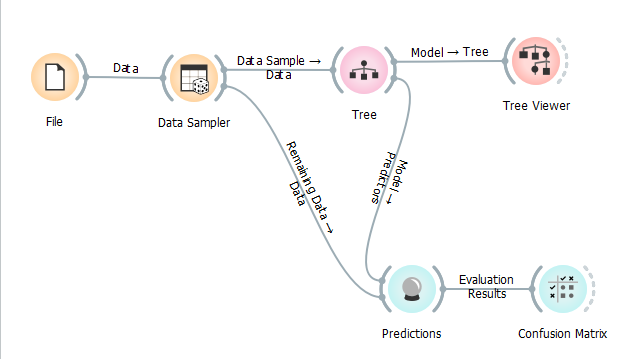
* 1 – INTRODUÇÃO
  1. – Contextualizando o Problema:

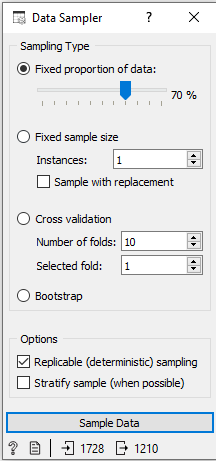
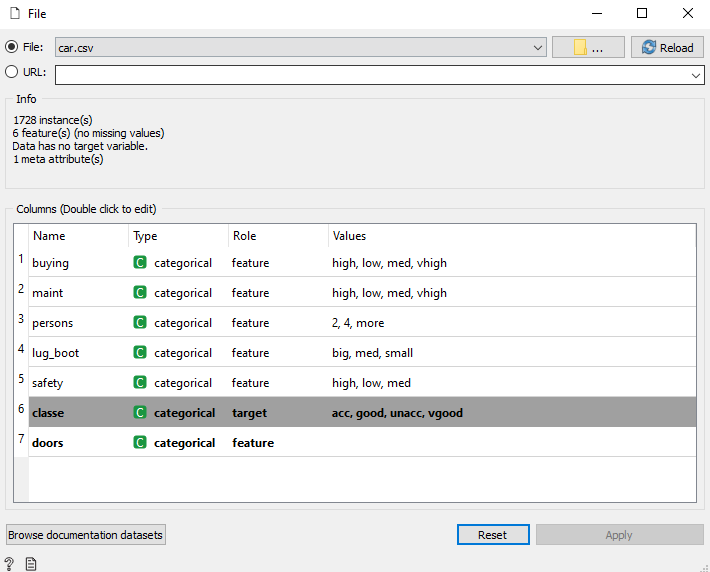
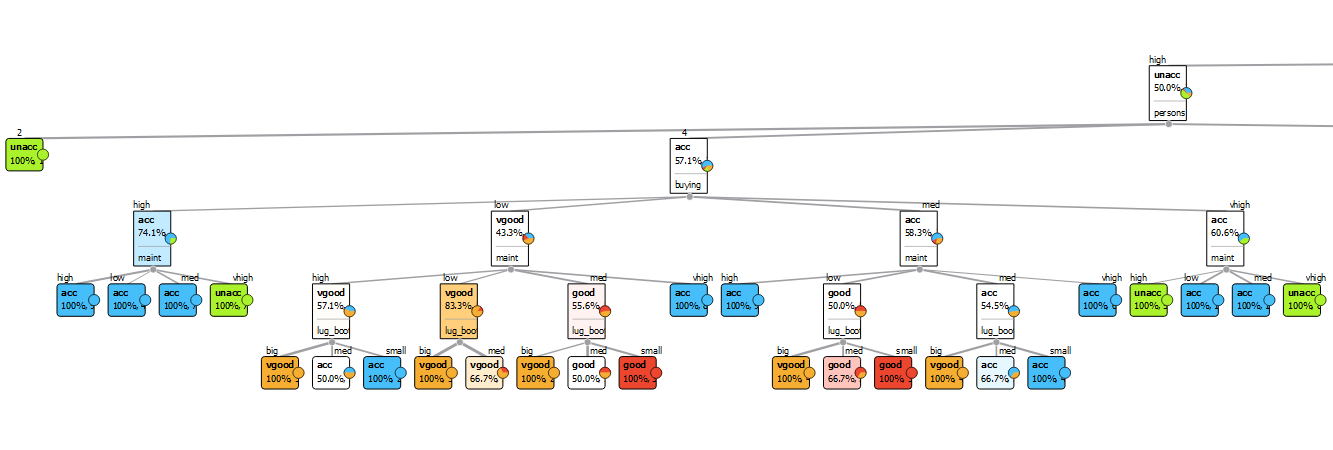
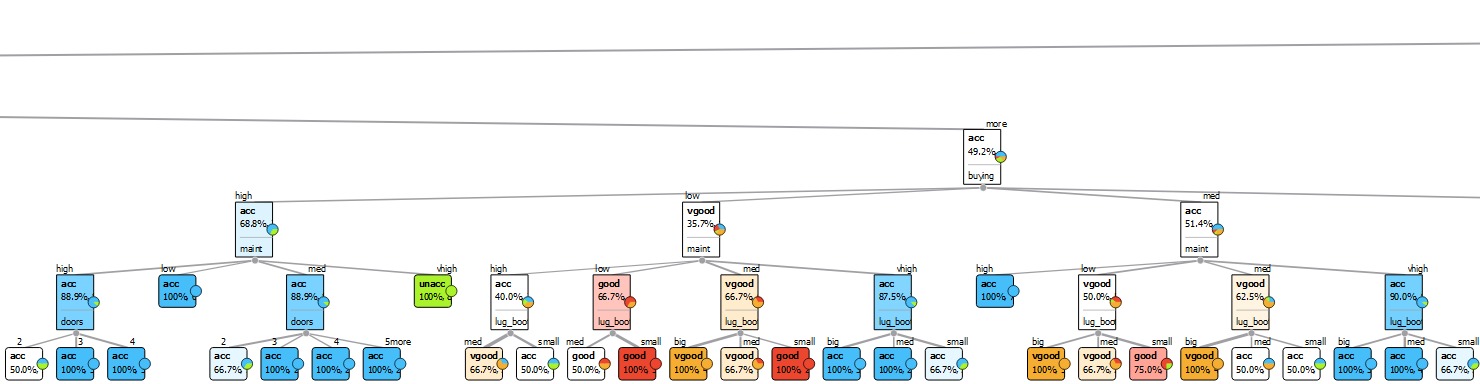
Neste projeto, utilizamos uma base de dados disponível na UCI Datasets, chamada Car Evaluation Data Set. Sendo assim, um data set com o intuito de classificar um determinado carro de acordo com as suas características, na qual estão sob análise nesta base de dados: Valor de compra, manutenção, portas, quantidade máxima de pessoas, tamanho do porta-malas e segurança.

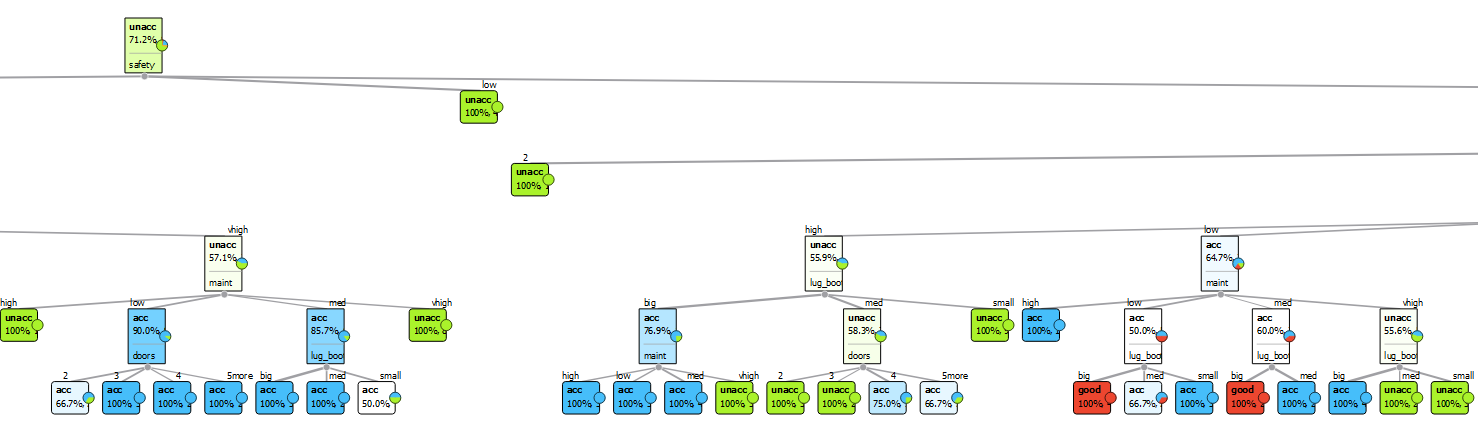
* 1. – Objetivo:

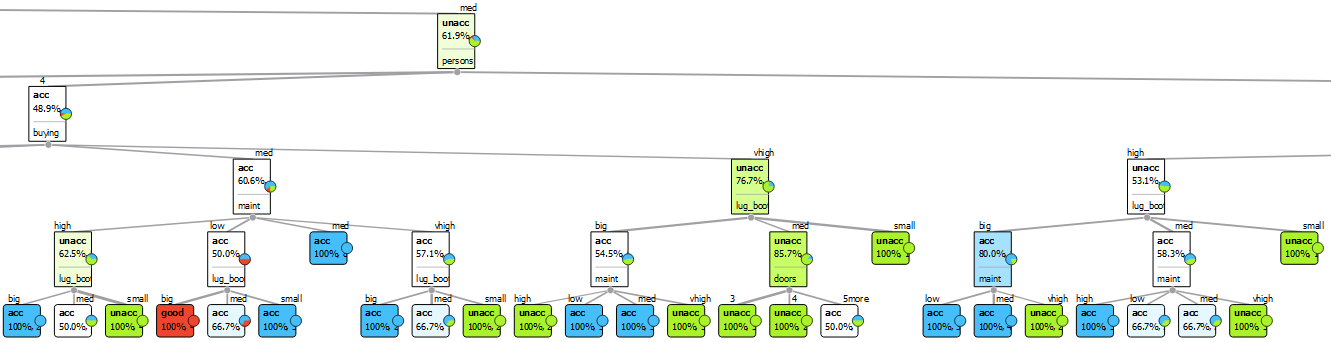
O objetivo é obter, através de perguntas simples, a classificação de um determinado carro, em que é considerado todos os detalhes e especificações na sua avaliação.

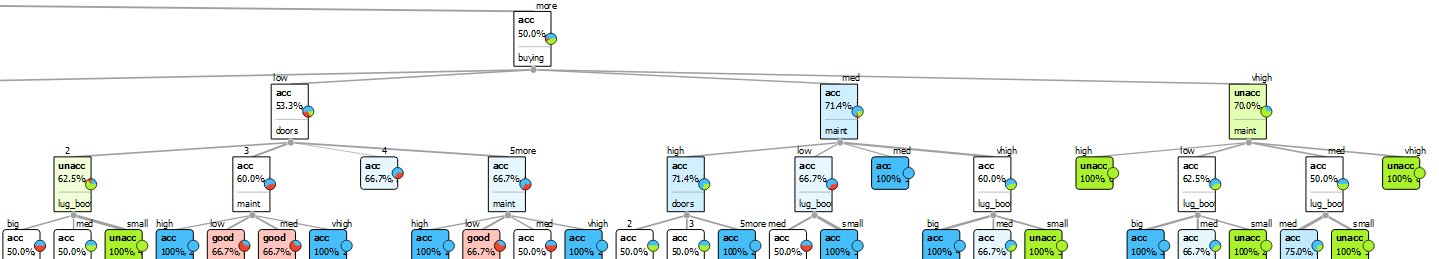
* 2 – ÁRVORE DE DECISÃO
  1. – Procedimentos:

 Para a realização da Árvore de Decisão, utilizamos a plataforma Orange, na qual a partir dela e da base de dados, foi possível gerar o esquema e posteriormente a matriz de confusão. Em seguida, carregamos a base de dados pra análise, e posteriormente separamos 70% de dados para análise e 30% de dados para teste. Com a parte de dados para análise, geramos a árvore de decisão e visualizamos. Com a parte de dados para teste, fizemos uma previsão, junto com os dados da árvore gerada, para finalmente gerar a matriz de confusão, na qual mostra os dados relacionados aos acertos feitos pela árvore.





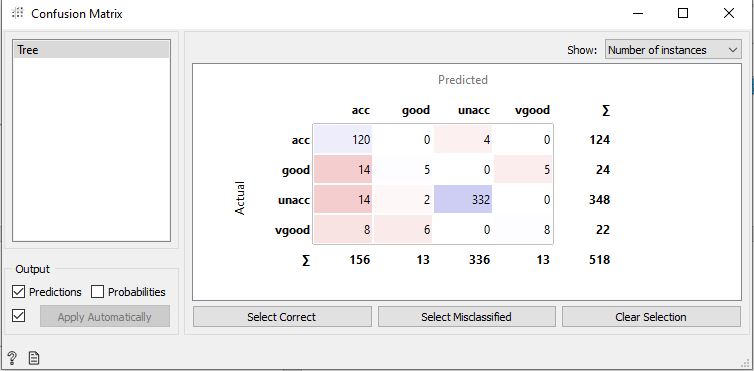




* 1. – Matriz de Confusão:

A matriz de confusão é calculada pela função de pesquisa Classificação, em que exibe a distribuição dos registros de acordo com suas classes atuais e as classes previstas. Sendo assim, os números fornecidos indicam a qualidade do modelo, tendo em vista a taxa de acerto.

Segue a seguir, a matriz do modelo utilizado:



Em uma Matriz de Confusão, é levado em consideração as linhas diagonais, na qual faz relação entre a classe prevista e a atual, sendo que na diagonal representam a mesma classe.

Neste modelo, conseguimos tirar algumas conclusões:

**-** A classe **Aceitável (acc)** foi a que teve maior taxa de acerto, tendo apenas 4 erros, classificados como **inaceitáveis (unacc).**

**-** A classe **Bom (good)** não obteve uma boa taxa de acerto, sendo que as classificações atuais, em sua maioria, indicaram **aceitáveis**, não condizente com as previsões.

**-** A classe **Inaceitável** obteve uma boa taxa de acerto, contudo houve algumas classificações erradas, apontando para aceitável e bom.

**-** A classe **Muito Bom (v\_good)** foi a classe que menos apareceu, em que nela a taxa de acerto não foi muito boa, sendo que houve empate entre as classificações muito bom e aceitável, além de obter classificação boa em determinados casos.

- Em geral, apenas as classificações Aceitáveis e Inaceitáveis obtiveram bons resultados no que diz respeito a assertividade. A classificação Boa e Muito Boa variou em seus resultados e não geraram resultados precisos, sendo assim, gerando conflito nos resultados gerados na Árvore de Decisão. Além disso, é notável a quantidade superior de classificações aceitáveis e inaceitáveis em relação as classificações boas e muito boas, tendo em vista a quantidade de vezes em que apareceram.

* 3 – APLICAÇÃO

Para fazer a aplicação desta Árvore de Decisão, desenvolvemos um programa em C++, no qual realizada perguntas simples relacionadas às classificações dos automóveis, sendo que internamente do programa foi realizado todas as variáveis e possibilidades possíveis que estão contidas na árvore de decisão. Dessa maneira, visto que existem muitas possibilidades distintas, fizemos todo o algoritmo usando “Switch”, variando para cada escolha feita.

Segue abaixo, parte do código feito:



