

## Bloco 3 \ Aulas

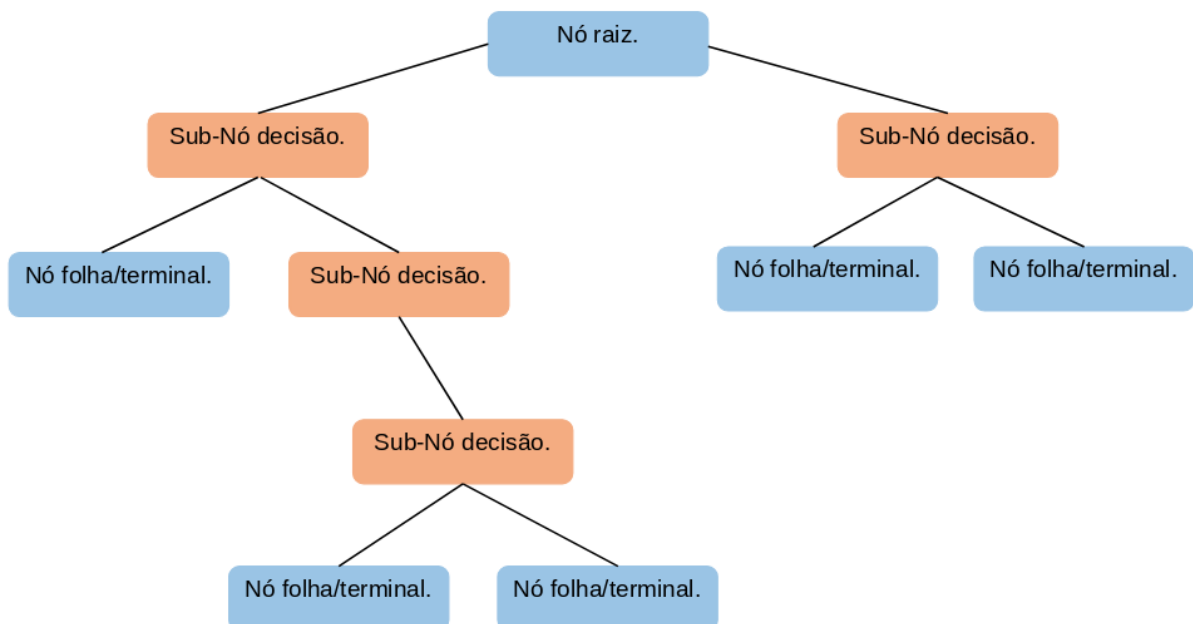
### 1 - Árvore de decisão.

Basicamente é um algoritmo de aprendizagem de máquina supervisionado utilizado nas tarefas de classificação e tarefas de regressão.

#### 1.1 - Objetivo.

Criar um modelo que preveja um valor de uma variável alvo aprendendo regras de decisão simples inferidas nos atributos dos dados.

#### 1.2 - Estrutura.



Fonte: (Autoria própria).

#### 1.3 - Conceitos das partes da estrutura:

**Nó raiz:** Uma das variáveis da base de dados, divide e melhora o conjunto de dados.

**Sub-nó de decisão:** Representa as condições If e else ou “se” e “se não”.

**Nó folha/ terminal:** Representa classe no caso da classificação, ou valor no caso da regressão.

#### 1.4 - Quais são as noções de uma árvore de decisão:

O objetivo de qualquer árvore de decisão é criar um modelo viável que preveja o valor de uma variável de destino com base no conjunto de variáveis de entrada (BELL, 2014).

## 1.5 Orientações da aplicação ORANGE:

Orange:

### Parte 1 inserir o “dataset”

**Primeiro passo abrir a aba Data:**

**Segundo passo:** arrastar “dataset” para o espaço de trabalho:

**Terceiro passo:** dar um duplo click sobre o "data set" assim abrindo a base de dados

**Quarto passo:** escolher a base de dados zoo.

### Parte 2 inserir o “data table”

Obs: Este exemplo de algoritmo destaca o tipo de classificação da árvore de decisão.

Depois de inserir o "data set" e selecionar a base de dados desejada feche a aba da base de dados e insira o "data table", depois é só criar uma conexão entre o "data set" e o “data table”, dessa forma com apenas um duplo click pode- visualizar os dados do “data table”.

### Quais são os dados do “data set” zoo?

As informações que estão armazenadas no “data set” zoo basicamente são informações de registro de vários animais de um zoológico que para simplificar são classificados pelo seu tipo por exemplo: mamíferos, anfíbios, aves, peixes, insetos, invertebrados, répteis. A partir desses dados também é exibido as características e o nome de cada tipo do animal que está presente na base de dados.

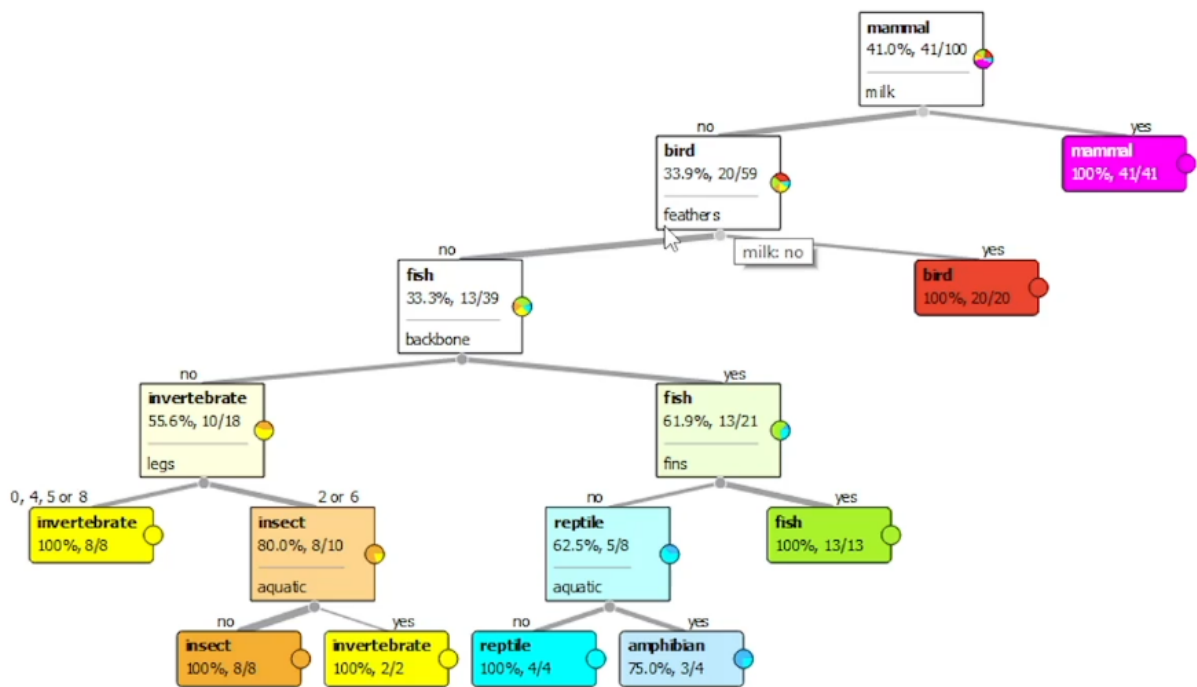
| Type      | Name    | Hair | Feathers | Eggs |
|-----------|---------|------|----------|------|
| Mammal    | Buffalo | Yes  | No       | No   |
| Amphibian | Frog    | No   | No       | Yes  |
| Bird      | Crow    | No   | Yes      | Yes  |
| Fish      | Carp    | No   | No       | Yes  |

### Parte 3 inserir o “Tree”.

Depois de visualizar os dados para aplicar o algoritmo neles é só selecionar o ícone tree da aba model e arrastar o ícone selecionado para a área de trabalho do “Orange” e depois é só arrastar até o “data set “ assim fazendo uma conexão entre eles e assim é possível a visualização do que seria uma parte da nossa árvore de decisão,

#### Parte 4 inserir o “Tree Viewer”.

Selecione o ícone “Tree Viewer” que está na aba “visualize”, crie uma conexão entre ele e o “tree” que no caso é o nosso algoritmo, assim dando dois clicks sobre o “Tree Viewer” ele mostra seus dados por exemplo: se um animal gera leite ele já pode ser classificado como mamífero se não, ele parte para a próxima decisão aqui os dados chegaram em um ponto muito importante como já foi citado os nossos If e else, dessa forma pode-se dizer que surgiu uma base do conceito da árvore de decisão como é mostrado na imagem a seguir:



Fonte().

**Decisão 1:** Se o animal produz leite ele é um mamífero. Se não, cai na condição 2.

**Decisão 2:** Se o animal possui penas ele é uma ave. Se não, cai na condição 3.

**Decisão 3:** Se o animal não possui coluna vertebral ele cai em um sub nó que se ele possui barbatanas é considerado um peixe se não o algoritmo irá avaliar se ele é aquático ou não se ele for aquático ele é classificado como anfíbio se não ele é um réptil.

**Decisão 4:** Se o animal possui coluna vertebral se ele possuir 0, 4, 5, 8 pernas ele é um animal invertebrado se não, se ele tiver 2 ou 6 pernas e for um animal aquático ele ainda é considerado invertebrado, mais se ele não for aquático é classificado como inseto.

#### Parte 4 “Test and Score”.

Arraste o “Test and Score” que está na aba “Evaluate” para a área de trabalho do “Orange” e crie uma conexão com o “Tree” e o “Data set” que é o nosso algoritmo, basta dar um

duplo click que o algoritmo mostrará a porcentagem que no caso exibirá a aproximação de 94%.

### **Testes algoritmos:**

Conectando os três algoritmos abaixo ao “Test and Score”, irá aparecer os resultados no caso as porcentagens como irá ser exibido a seguir:

#### **Neural Network:**

O algoritmo com a maior acurácia ou porcentagem com 96%.

#### **Tree:**

O algoritmo com a segunda melhor acurácia ou porcentagem de 94%.

#### **KNN:**

O algoritmo com a terceira melhor acurácia ou porcentagem de 91%.

#### **Naive Bayes:**

O algoritmo com a menor acurácia ou porcentagem 90%.