|  |  |
| --- | --- |
|  | **UNOESTE**  **SISTEMAS DE INFORMAÇÃO** |

FELIPE AUGUSTO GOBBO DOS SANTOS: RA: 262319020

LUIGI SARDELARI SCALIANTE: RA: 262318687

JOÃO CLÁUDIO MARTINEZ DA COSTA: RA:262319640

BHRENO GABRIEL MORAES FERREIRA: RA: 262318792

**APLICAÇÃO COM ÁRVORE DE DECISÃO**

Presidente Prudente – SP

2023

**Sumário**

**•** [**A BASE DE DADOS: 3**](#_Toc152083542)

**•** [**Introdução 3**](#_Toc152083543)

**•** [**SOBRE A BASE DE DADOS: 3**](#_Toc152083544)

**•** [**TIPOS DE ANIMAIS 4**](#_Toc152083545)

**•** [**Mamíferos Terrestres 4**](#_Toc152083546)

**•** [**Aves 4**](#_Toc152083547)

**•** [**Répteis 4**](#_Toc152083548)

**•** [**Peixes 4**](#_Toc152083549)

**•** [**Anfíbios 4**](#_Toc152083550)

**•** [**Insetos 4**](#_Toc152083551)

**•** [**Moluscos e Artrópodes Aquáticos 4**](#_Toc152083552)

**•** [**ATRIBUTOS 5**](#_Toc152083553)

**•** [**ÁRVORE DE DECISÃO E MATRIZ DE CONFUSÃO 6**](#_Toc152083554)

**•** [**ÁRVORE DE DECISÃO 6**](#_Toc152083555)

**•** [**PREDIÇÕES 7**](#_Toc152083556)

**•** [**ÁNALISE DE RESULTADOS: 8**](#_Toc152083557)

**•** [**AUC (Area Under The Curve): 8**](#_Toc152083558)

**•** [**CA (Classification Accuracy) 8**](#_Toc152083559)

**•** [**Precision 8**](#_Toc152083560)

**•** [**Recall 8**](#_Toc152083561)

**•** [**F1 8**](#_Toc152083562)

**•** [**Execução do Programa 9**](#_Toc152083563)

# A BASE DE DADOS:

## Introdução

Explicação breve sobre o assunto e o tipo de linguagem usada:

O tema escolhido para a criação da árvore e desenvolvimento do código foi animais do zoológico. Na qual consiste em chegar ao resultado correto (espécie e nome do animal) através das características de cada um. Sendo assim com base na estrutura de árvore (algoritmo pronto resultado pelo Orange) inicializamos o algoritmo com linguagem C.

.

# 

# SOBRE A BASE DE DADOS:

A base de dados contém perguntas relacionadas a características de animais, gerais de animais, sejam físicas, tipo de pelagem, tamanho, domesticidade... E com base nos resultados, podemos chegar à sua classificação específica.

Link para a base de dados:

<https://archive.ics.uci.edu/dataset/111/zoo>

# TIPOS DE ANIMAIS

## Mamíferos Terrestres

Porco-da-terra, Antílope, Urso, Javali, Búfalo, Bezerro, Preá, Guepardo, Veado, Elefante, Morcego-da-fruta, Girafa, Cabra, Gorila, Hamster, Lebre, Leopardo, Leão, Lince, Vison, Toupeira, Mangustoo, Gambá, Órix, Ornitorrinco, Doninha, Pônei, Toninha, Puma, Gato, Guaxinim, Rena, Foca, Leão Marinho, Esquilo, Morcego-Vampiro, Ratazana, Canguru, Lobo.

## Aves

Galinha, Corvo, Pomba, Pato, Flamingo, Gaivota, Falcão, Kiwi, Alaudidae, Avestruz, Periquito, Pinguim, Faisão, Réia, Talha-Mar, Skua, Pardal, Cisne, Abutre, Carriça.

## Répteis

Víbora, Cobra d’água, Verme-lento, Tartaruga, Tuatara.

## Peixes

Perca, Carpa, Peixe-Gato, Chub, Tubarão, Arinca, Herring, Lúcio, Piranha, Cavalo-Marinho, Solha, Arraia, Atum.

## Anfíbios

Sapo, Salamandra, Rã, Perereca.

## Insetos

Pulga, Mosquito, Abelha, Mosca, Joaninha, Mariposa, Cupim, Vespa.

## Moluscos e Artrópodes Aquáticos

Molusco, Carangueijo, Lagostim, Lagosta, Pouvo, Escorpião, Vespa-do-mar, Lesma, Estrela-do-mar, Minhoca.

# ATRIBUTOS

(Nome do atributo e tipo de domínio de valor).

1. nome do animal: Exclusivo para cada valor

2. pelo Boolean

3. penas Boolean

4. ovos Boolean

5. leite Boolean

6. aéreo Boolean

7. aquático Boolean

8. predador Boolean

9. dentes Boolean

10. coluna vertebral Boolean

11. respiração Boolean

12. venenoso Boolean

13. barbatanas Boolean

14. pernas Numérico (conjunto dos valores: {0,2,4,5,6,8})

15. cauda Boolean

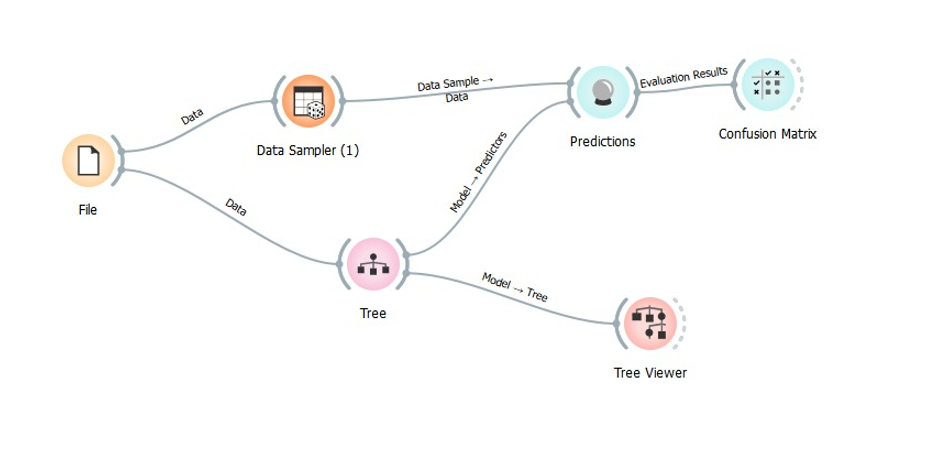
16. doméstico Boolean

17. tamanho do gato Boolean

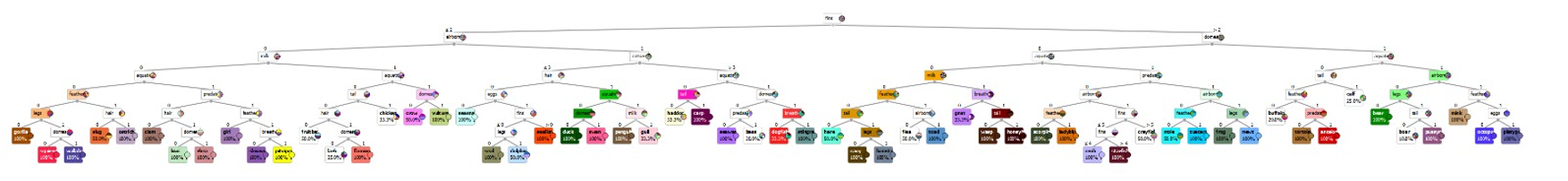
18. tipo Numérico (valores inteiros no intervalo [1,7])

# ÁRVORE DE DECISÃO E MATRIZ DE CONFUSÃO

Para gerarmos a Árvore de Decisão e a Matriz de Confusão, montamos a seguinte estrutura no programa Orange:



## ÁRVORE DE DECISÃO

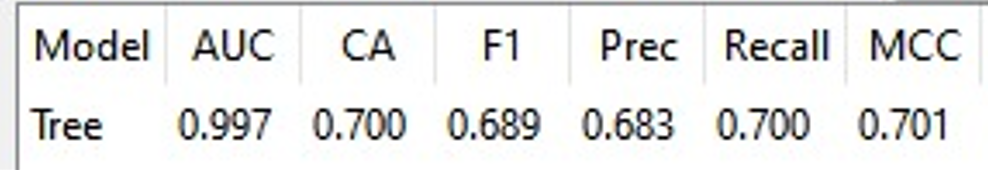
****

(Devido ao tamanho, a árvore ficou distorcida no word)

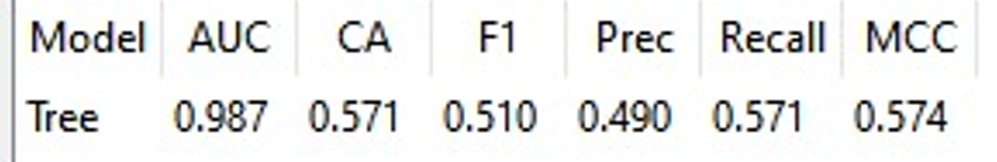
## PREDIÇÕES

A árvore foi gerada com um conjunto de dados de 70% para treino e 30% para teste. Aqui estão os dados:

30%



70%

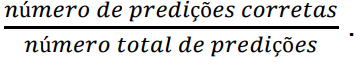


# ÁNALISE DE RESULTADOS:

AUC (Area Under The Curve): indica o quão exato o modelo é em sua classificação. Ou seja, o quão preciso é ao indicar a classe de um registro.

Quanto mais próximo de 1, melhor: nosso modelo alcançou uma pontuação de 0,990, indicando que é bastante preciso ao indicar o nível de obesidade de um registro.

CA (Classification Accuracy)**:** a acurácia do modelo é dada pela fórmula 𝐶𝐴 =

.

Nosso modelo apresentou um CA de 0,590, o que não é ideal, pois o valor máximo e

Melhor é 1.

Precision: Semelhante a CA, a Precision mede quantas vezes o modelo acertou em relação ao total de vezes em que ele tenta acertar.

Recall: Indica a quantidade de vezes em que o modelo acertou a classe em relação

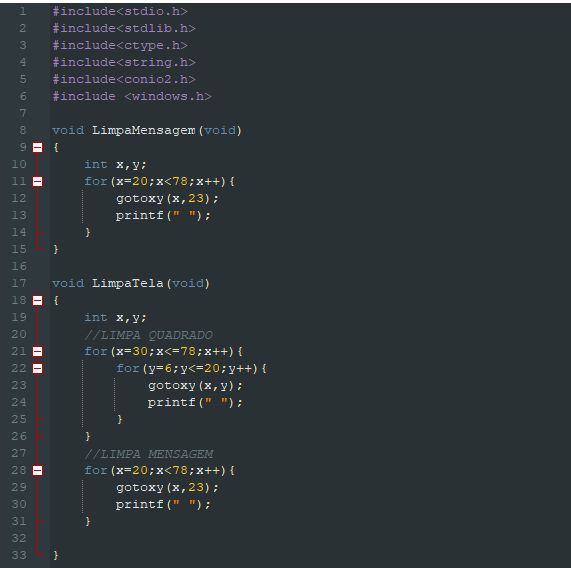
À quantidade de vezes em que deveria ter acertado.

F1**:** Indica o quão Precision e Recall estão condizentes entre si. Ou seja, quanto maior for F1, melhor, pois Precision e Recall estão em níveis parecidos (não

Necessariamente alto ou baixo, mas, sim, próximos).

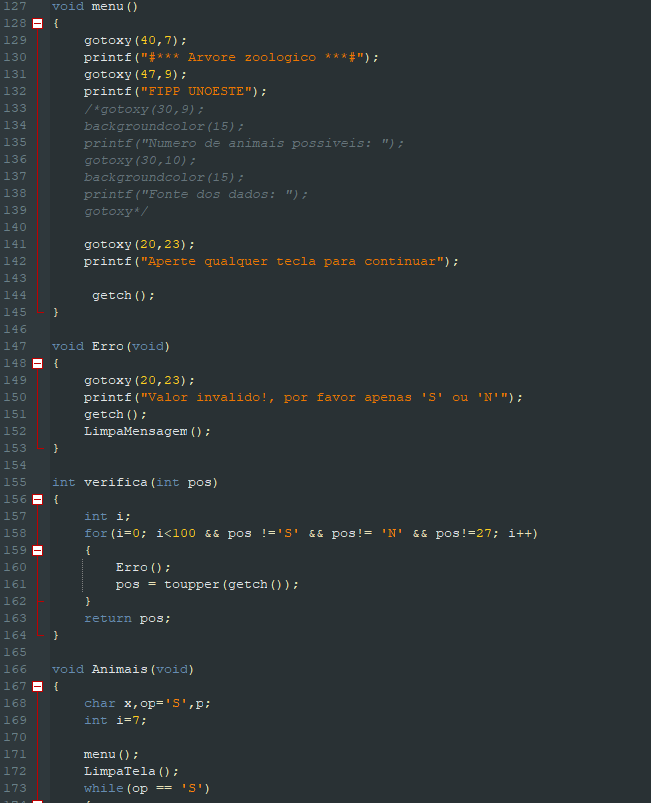
# Execução do Programa

Para melhor entender a dinâmica do programa mostraremos o início:



Acima na imagem é mostrado o começo do código, importamos bibliotecas e criamos funções para desenvolver.

Logo após os créditos e as criações, começamos as perguntas:



Finalizando a primeira parte do código que inicializa o programa, começamos as decisões (if else):



Assim foi se desenvolvendo o programa na temática “Zoo”.

E ao fim do programa temos as devidas decisões tomadas de forma correta conforme as características do animal:

