--



**TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN - TP HCM**

**KHOA CÔNG NGHỆ THÔNG TIN**

**MÔN CƠ SỞ TRÍ TUỆ NHÂN TẠO**

**LỚP CỬ NHÂN TÀI NĂNG 2016**

Machine Learning

ĐOÀN QUANG TUẤN 1612780

**LÊ HOÀNG SANG 1612554**

NHÓM SINH VIÊN THỰC HIỆN

**BÁO CÁO ĐỒ ÁN**

1. **Tìm hiểu công cụ Weka.**
2. **Sử dụng Weka để chạy thuật toán ID3.**
3. **Mô tả dữ liệu Zoo.**

* Số mẫu trong tập dữ liệu là 101 loại động vật.
* Tên và ý nghĩa các thuộc tính (18):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **STT** | **Thuộc tính** | **Kiểu dữ liệu** | **Ý nghĩa** | **Miền giá trị** |
| 1 | name | String | Tên động vật | Chuỗi |
| 2 | hair | Boolean | Tóc | 1,0 |
| 3 | feathers | Boolean | Lông | 1,0 |
| 4 | eggs | Boolean | Đẻ trứng | 1,0 |
| 5 | milk | Boolean | Có sữa | 1,0 |
| 6 | airbone | Boolean | Biết bay | 1,0 |
| 7 | aquatic | Boolean | Đông vật sống dưới nước | 1,0 |
| 8 | predator | Boolean | Động vật ăn thịt | 1,0 |
| 9 | toothed | Boolean | Có răng | 1,0 |
| 10 | backbone | Boolean | Có xương sống | 1,0 |
| 11 | breathes | Boolean | Có thở | 1,0 |
| 12 | venomous | Boolean | Có độc | 1,0 |
| 13 | fins | Boolean | Có vây | 1,0 |
| 14 | legs | Numeric | Số chân | 0,2,4,5,6,8 |
| 15 | tail | Boolean | Có đuôi | 1.0 |
| 16 | domestic | Boolean | Động vật được nuôi | 1,0 |
| 17 | type | Numeric | Lớp động vật | 1,2,3,4,5,6,7 |

* Danh sách phân lớp, đặt tên cho mỗi phân lớp:

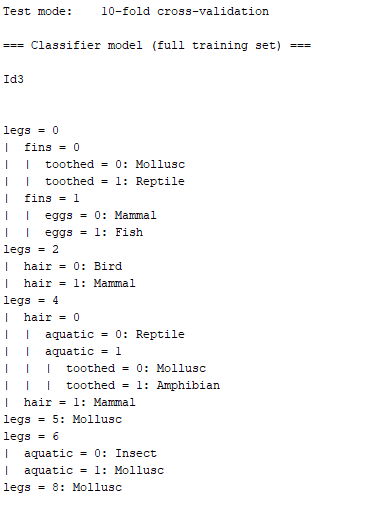
|  |  |  |  |
| --- | --- | --- | --- |
| **Type** | **Đặt tên** | **Số lượng** | **Động vật thuộc lớp** |
| 1 | Mammal | 44 | aardvark, antelope, bear, boar, buffalo, calf, cavy, cheetah, deer, dolphin, elephant, fruitbat, giraffe, girl, goat, gorilla, hamster, hare, leopard, lion, lynx, mink, mole, mongoose, opossum, oryx, platypus, polecat, pony, porpoise, puma, pussycat, raccoon, reindeer, seal, sealion, squirrel, vampire, vole, wallaby,wolf |
| 2 | Bird | 20 | chicken, crow, dove, duck, flamingo, gull, hawk, kiwi, lark, ostrich, parakeet, penguin, pheasant, rhea, skimmer, skua, sparrow, swan, vulture, wren. |
| 3 | Reptile | 5 | pitviper, seasnake, slowworm, tortoise, tuatara. |
| 4 | Fish | 13 | bass, carp, catfish, chub, dogfish, haddock, herring, pike, piranha, seahorse, sole, stingray, tuna. |
| 5 | Amphibian | 3 | frog, newt, toad |
| 6 | Insect | 8 | flea, gnat, honeybee, housefly, ladybird, moth, termite, wasp. |
| 7 | Mollusc | 10 | clam, crab, crayfish, lobster, octopus, scorpion, seawasp, slug, starfish, worm |

1. **Sử dụng thuật toán ID3 để học ra cây quyết định.**

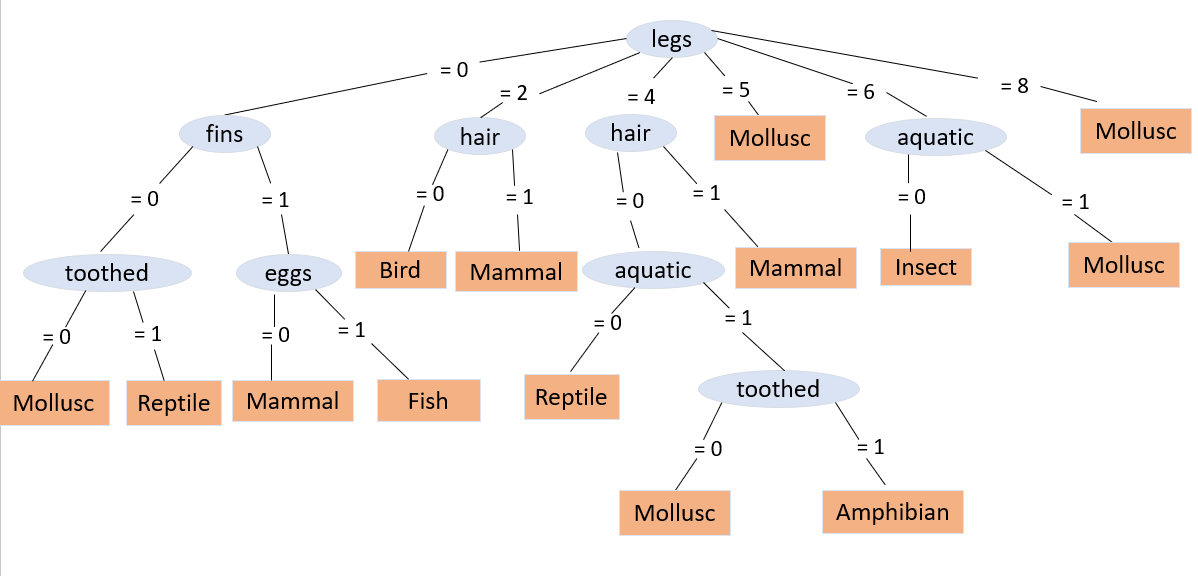
Sử dụng thuật toán id3 cho thuộc tính type. Tên của động vật chúng ta không cần xét đến nên ta sử dụng bộ dữ liệu không có thuộc tính name: **Zoo-train.arff**

- Test option: 10-fold cross validation

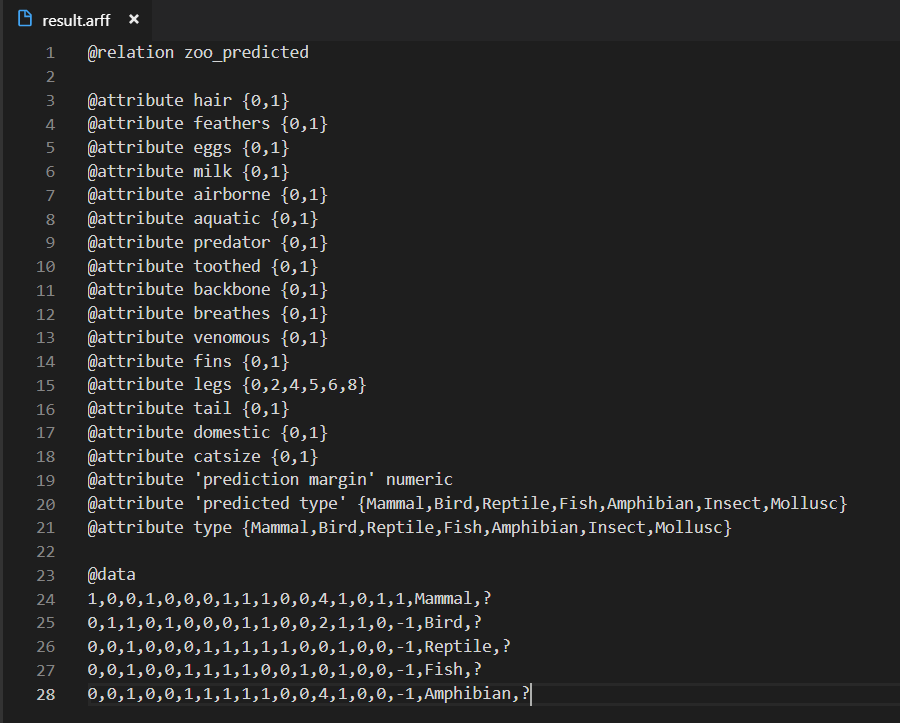
Ta được cây quyết định:



Cây được vẽ lại:



* Kết quả 5 mẫu đã cho dự đoán từ cây quyết định đã sinh ra là:



1. Mamal
2. Bird
3. Reptile
4. Fish
5. Amphibian
6. **Chạy các thuật toán khác.**
7. **Chương trình python cho giải thuật Naïve Bayes**