

**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 100 (mẫu).
* Tên và ý nghĩa các thuộc tính:
* Thuộc tính name: thuộc tính kiểu định danh, tên là duy nhất cho mỗi đối tượng. 100 mẫu tương ứng 100 tên.
* Thuộc tính hair: thuộc tính kiểu boolean, xác định đối tượng có tóc hay là không. (1: có, 0: không).
* Thuộc tính feathers: thuộc tính kiểu boolean, xác định đối tượng có lông hay là không. (1: có, 0: không).
* Thuộc tính eggs: thuộc tính kiểu boolean, xác định đối tượng đẻ trứng hay đẻ con (1: đẻ trứng, 0: đẻ con).
* Thuộc tính milk: thuộc tính kiểu boolean, xác định đối tượng có sữa hay không (nuôi con bằng sữa không). (1: có, 0: không).
* Thuộc tính airborne: thuộc tính kiểu boolean, xác định đối tượng biết bay hay không. (1: có, 0: không).
* Thuộc tính predator: thuộc tính kiểu boolean, xác định đối tượng phải là động vật ăn thịt hay không. (1: phải, 0: không).
* Thuộc tính airborne: thuộc tính kiểu boolean, xác định đối tượng biết bay hay không. (1: có, 0:không).
* Thuộc tính toothed: thuộc tính kiểu boolean, xác định đối tượng là động vật có răng không. (1: có, 0:không).
* Thuộc tính backbone: thuộc tính kiểu boolean, xác định đối tượng là động vật có xương sống hay không. (1: có, 0:không).
* Thuộc tính breathes: thuộc tính kiểu boolean, xác định đối tượng là động vật có thở không. (1: có, 0:không).
* Thuộc tính venomous: thuộc tính kiểu boolean, xác định đối tượng là động vật có độc không. (1: có, 0:không).
* Thuộc tính fins: thuộc tính kiểu boolean, xác định đối tượng là động vật có vây không. (1: có, 0:không).
* Thuộc tính legs: thuộc tính kiểu số nguyên chẵn, biểu diễn số chân của động vật đó (0,2,4,6,8).
* Thuộc tính tail: thuộc tính kiểu boolean, xác định đối tượng là động vật có đuôi không. (1: có, 0:không).
* Thuộc tính domestic: thuộc tính kiểu boolean, xác định đối tượng là động vật được nuôi trong nhà không. (1: có, 0:không).
* Thuộc tính catsize: thuộc tính kiểu boolean, xác định đối tượng là động vật có … không. (1: có, 0:không).
* Thuộc tính type: thuộc tính kiểu số nguyên, phân lớp động vật - 7 lớp (1-7).
* Danh sách các phân lớp:
* 1 - Lớp thú (41): 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 - Lớp chim (20): chicken, crow, dove, duck, flamingo, gull, hawk, kiwi, lark, ostrich, parakeet, penguin, pheasant, rhea, skimmer, skua, sparrow, swan, vulture, wren.
* 3 - Lớp bò sát có đuôi (5): pitviper, seasnake, slowworm, tortoise, tuatara.
* 4 - Lớp cá (13): bass, carp, catfish, chub, dogfish, haddock, herring, pike, piranha, seahorse, sole, stingray, tuna.
* 5 - Lớp bò sát không đuôi (3): frog, newt, toad
* 6 - Lớp côn trùng (8): flea, gnat, honeybee, housefly, ladybird, moth, termite, wasp.
* 7 – Động vật sống dưới nước (10): clam, crab, crayfish, lobster, octopus, scorpion, seawasp, slug, starfish, worm.

1. **Sử dụng thuật toán ID3 để học ra cây quyết định.**
2. **Chạy các thuật toán khác.**
3. **Chương trình python cho giải thuật Naïve Bayes**