BỘ GIÁO DỤC VÀ ĐÀO TẠO

**TRƯỜNG ĐẠI HỌC CẦN THƠ**

**KHOA CÔNG NGHỆ THÔNG TIN & TRUYỀN THÔNG**

**🙞 🕮 🙜**



**MÁY HỌC NÂNG CAO**

**Đề tài:**

**ỨNG DỤNG DỰ ĐOÁN LOÀI ĐỘNG VẬT**

**Các sinh viên thực hiện:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Tên:** | **Đỗ Trung Nguyên** | **Đỗ Thành Công** | **Nguyễn Hưng** |
| **Mã số:** | **B1709552** | **B1709526** | **B1709536** |
| **Khóa:** | **43** | **43** | **43** |

Cần Thơ

NHẬN XÉT CỦA GIẢNG VIÊN

………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………

………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………

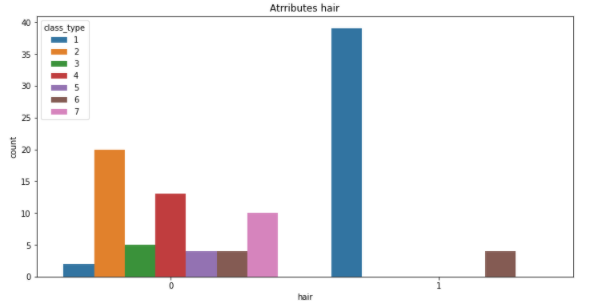
………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………

………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………

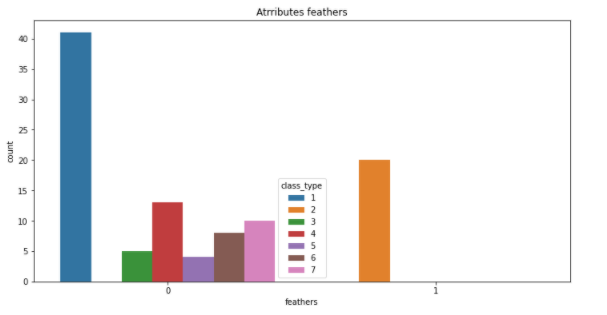
**CÁC CÔNG VIỆC ĐÃ THỰC HIỆN**

1. **Thu thập và giới thiệu tập dữ liệu**

* Dataset Zoo gồm 2 file
* zoo.csv
* class.csv
* File zoo.csv gồm 17 trường thuộc tính chứa các đặc điểm đế phân loại loài động vật và 1 cột nhãn chứa tên các loài.
  + Thuộc tính
    - animal\_name ( Tên động vật )
    - hair ( Lông, tóc )
    - feathers ( Lông vũ )
    - eggs ( Sinh sản bằng trứng )
    - milk (Có sữa )
    - airborne ( Môi trường trên không )
    - aquatic ( Môi trường dưới nước )
    - predator ( Động vật ăn thịt )
    - toothed ( Răng )
    - backbone ( Xương sống )
    - breathes ( Thở )
    - venomous ( Có nộc độc )
    - fins ( Vây )
    - legs ( Chân )
    - tail ( Đuôi )
    - domestic ( Bầy đàn )
    - Catsize ( Kích thước tương đối so với loài mèo )
  + Nhãn
    - class\_type
      * 1 : Động vật có vú
      * 2 : Chim
      * 3 : Bò sát
      * 4: Cá
      * 5 : Lưỡng cư
      * 6 : Bọ
      * 7 : Động vật không xương sống
* Vẽ một số biểu đồ tương quan giữa các thuộc tính với nhãn trong tập dữ liệu.
* **Lông (hair)**

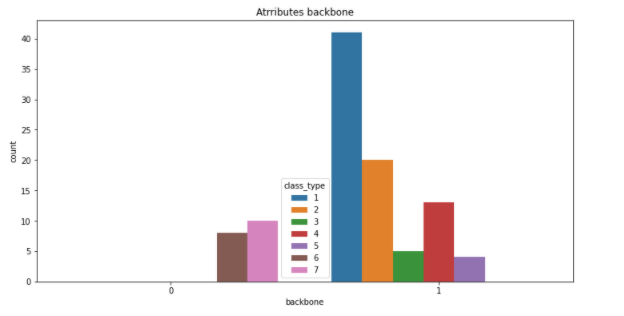


Các loài nằm trong khung tỉ lệ 0 và 1, đa số các loài thuộc hệ ‘động vật có vú’ thường có bộ lông và một số ít không có lông.

* **Lông vũ (feathers)**

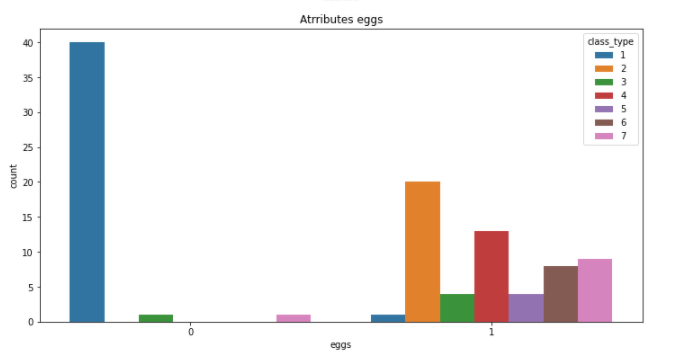
Theo biểu đồ thì các loài thuộc hệ 2 (chim) thì 100% đều có lông vũ.

* **Xương sống (backbone)**



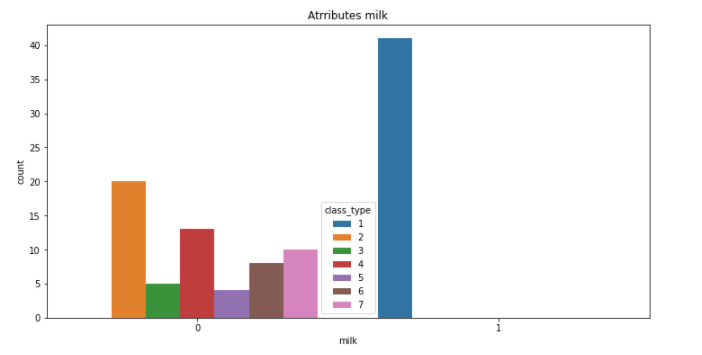
Về thuộc tính xương sống thì đa phần các loài trong nhãn đều có xương sống, ngoại trừ hệ 6 (bọ) và hệ 7 (không xương sống).

* **Trứng (eggs)**



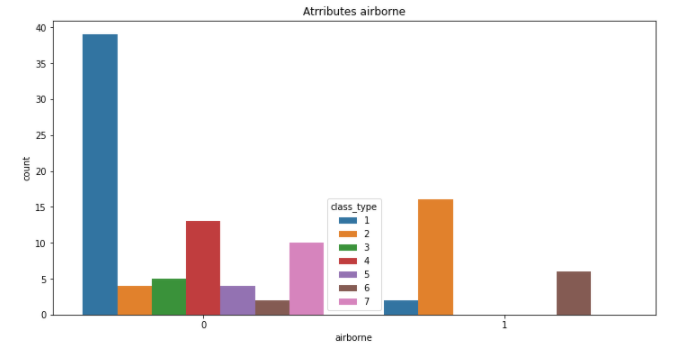
Đa phần các loài thuộc hệ 2 (chim), 3 (bò sát) và 4 (cá), thường đẻ trứng và một ít loài lưỡng cư.

* **Sữa (milk)**



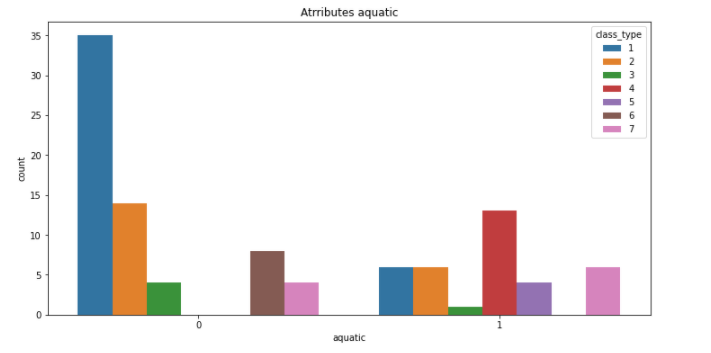
Biểu đồ chỉ rõ hệ 1 (động vật có vú) thì có sữa.

* **Môi trường trên không (airborne)**



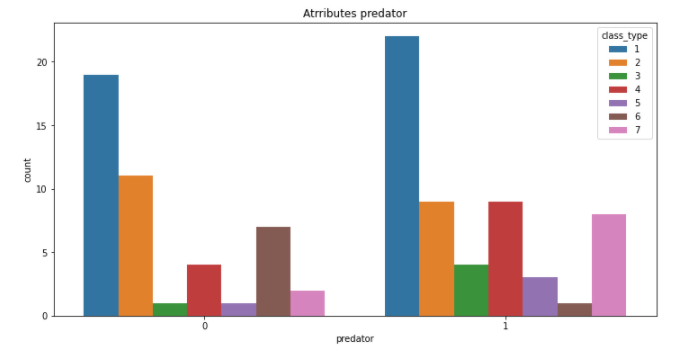
Hệ 2 (chim) là hệ tỉ lệ cao sống ở môi trường trên không và một ít loài thuộc hệ 6 (bọ).

* **Môi trường dưới nước (aquatic)**



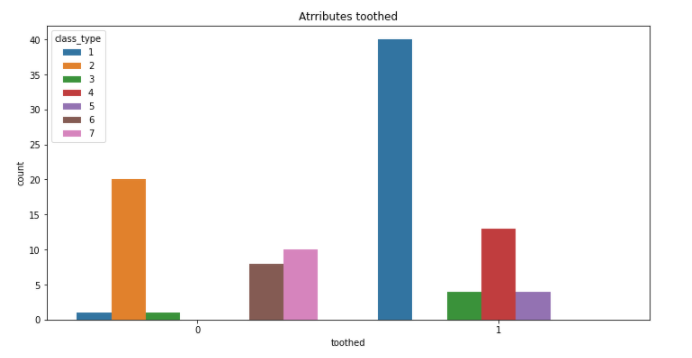
Một số loài trong các hệ 1 (động vật có vú), hệ 2 (chim), hệ 3 (bò sát) và phần lớn loài hệ 4 (cá) sinh sống và hoạt động được ở môi trường trong nước.

* **Động vật ăn thịt (predator)**



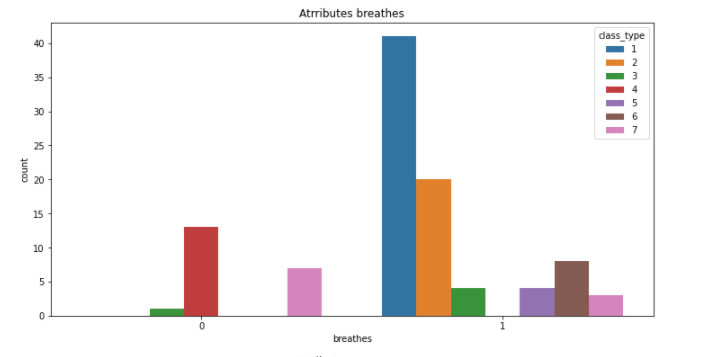
Các loài thuộc hệ 1 (động vật có vú), hệ 2 (chim) , hệ 3 (bò sát), hệ 4 (bò sát) và một ít loài hệ 5(lưỡng cư), hệ 6 (bọ) thường ăn thịt và cũng có một số loài không ăn thịt.

* **Có răng (toothed)**



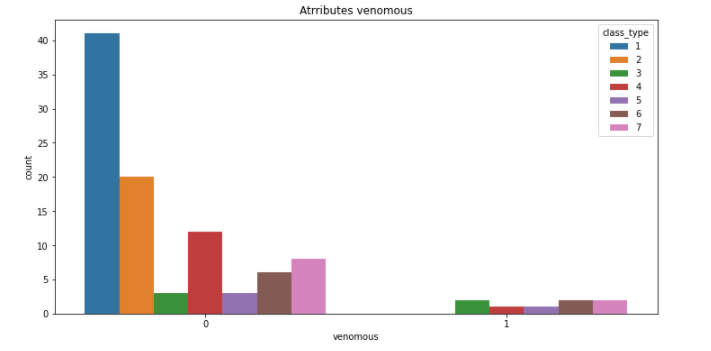
Hê 1 (động vật có vú) thường có răng và một ít loài thuộc hệ 3 (bò sát) và hệ 4 (cá) có răng.

* **Thở trên mặt đất (breathes)**



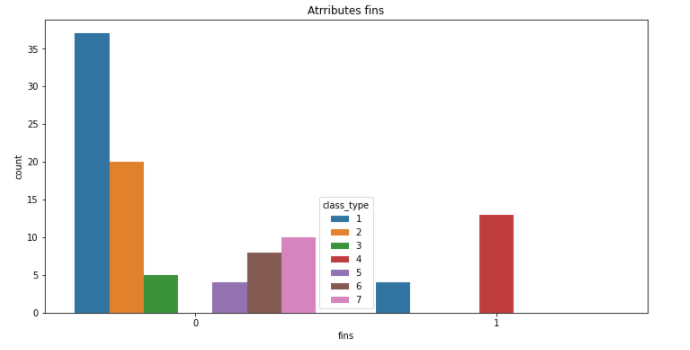
Đa số các loài hệ 4 (cá) không thể thở trên mặt đất.

* **Có nộc độc (venomous)**



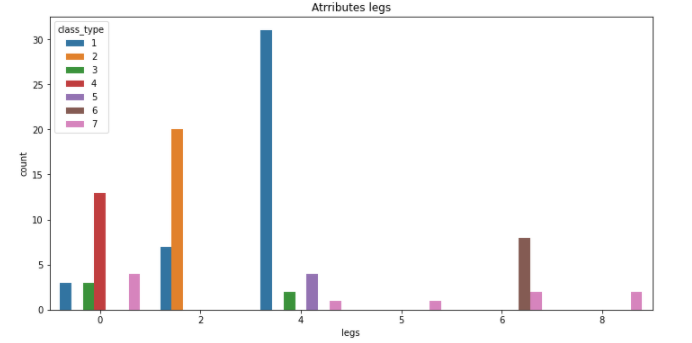
Một ít loài hệ 3 (cá), hệ 4 (bò sát), hệ 5 (lưỡng cư), hệ 6 (bọ) và hệ 7 (không xương sống) có nộc độc. Đa phần là không có.

* **Vây (fins)**



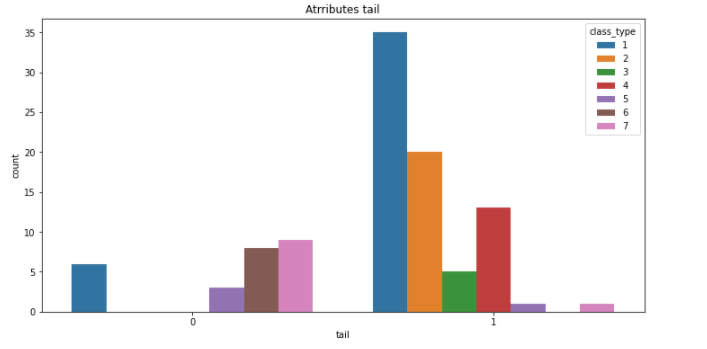
Các loài thuộc hệ 4 (cá) là có vây, các hệ khác hoàn toàn không có.

* **Chân (legs)**



Các loài trong tập dữ liệu có số lượng chân là 0,2,4,5,6,8. Phần lớn hệ 1 (động vật có vú) thuộc 4 chân, hệ 2 (chim) thuộc 2 chân.

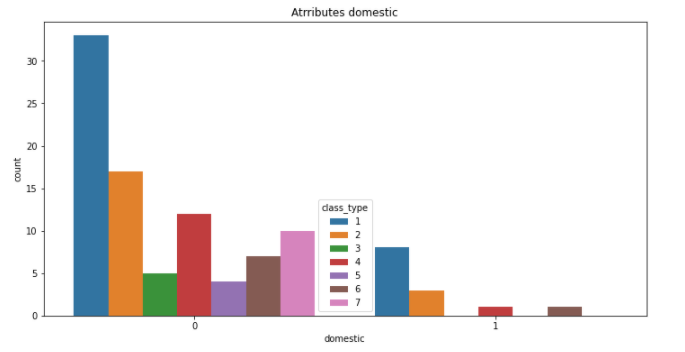
* **Đuôi (tail)**



Các loài trong các hệ trong nằm trong thuộc tính đuôi gồm:

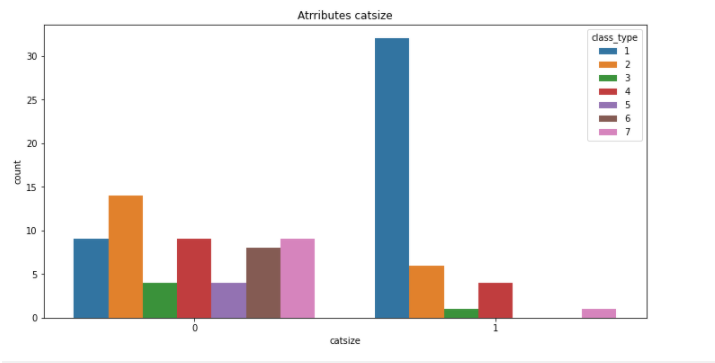
hệ 1 (động vật có vú), hệ 2 (chim) , hệ 3 (bò sát) , hệ 4 (cá).

* **Bầy đàn (domestic)**

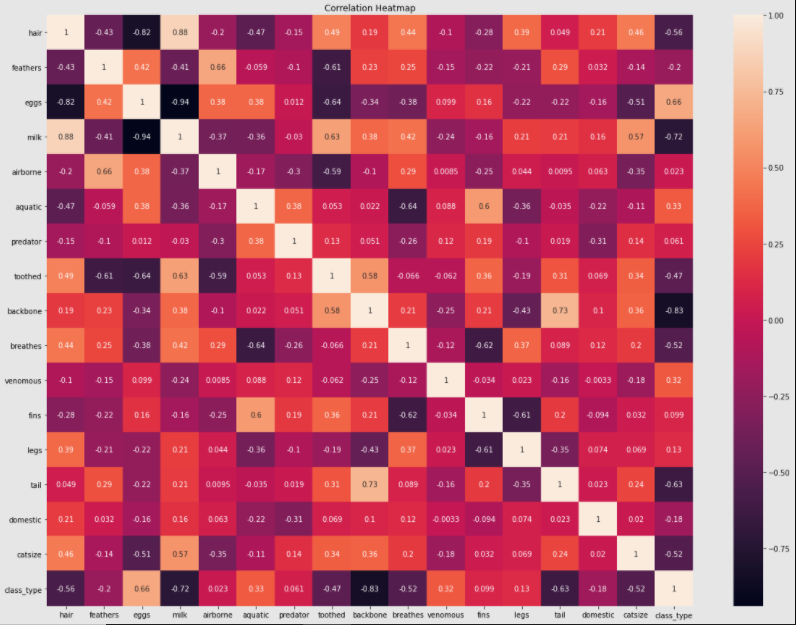


Theo biểu đồ nhận thấy chỉ một số loài trong hệ 1 (động vật có vú), hệ 2 (chim) và hệ 4 (cá) sẽ sống theo bầy.

* **Kích thước tương đối so với loài mèo (catsize)**



* **Vẽ biểu đồ nhiệt thể hiện giá trị tương quan giữa các thuộc tính trong hệ với nhau**.



Dựa trên biểu đồ ta thấy được độ tương quan giữa thuộc tính hair và milk là mật thiết với nhau (0.88).

1. **Huấn luyện mô hình**

* Sau khi biết được sự tương quan giữa thuộc tính với nhãn. Sử dụng thuật toán SVM để huấn luyện mô hình.
* Sử dụng các kernel khác nhau để so sánh kết quả.

*0* : linear

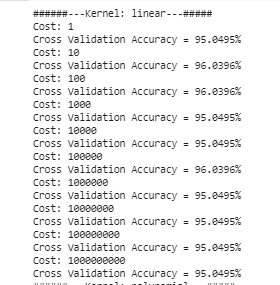
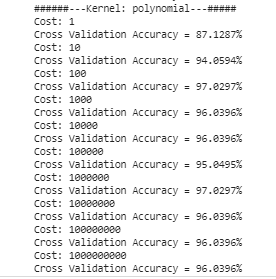
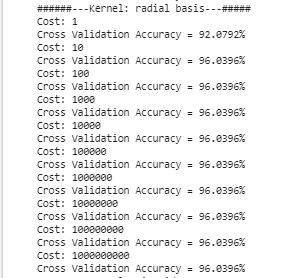
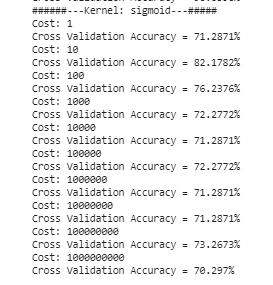
*1* : polynomial

*2* : radial basis

*3* : sigmoid

* Với mỗi kernel sử dụng các tham số c từ 1,10,100,…,1000000000. Để lựa chọn ra mô hình phù hợp nhất.

1. **Đánh giá mô hình**

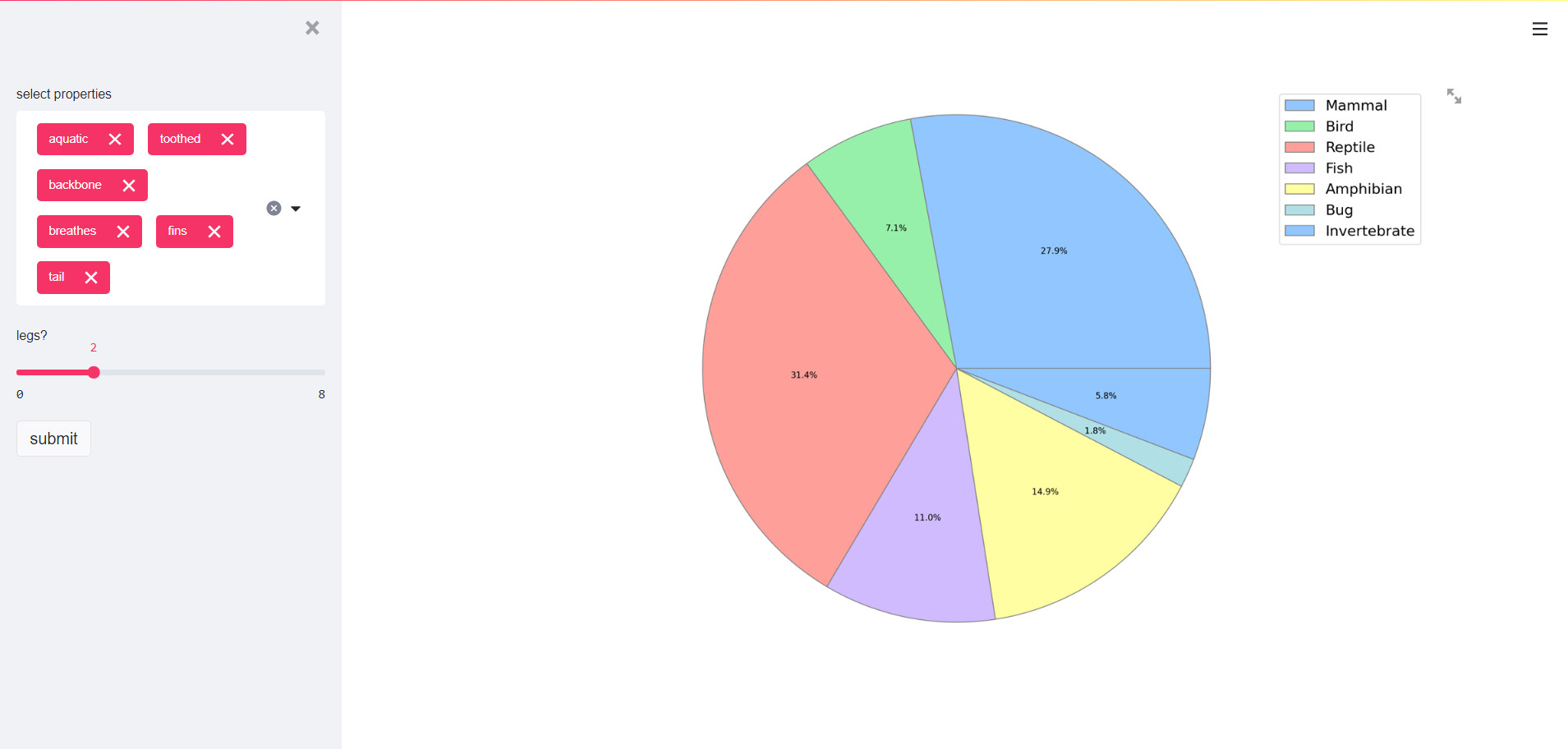
* Kết quả đánh giá dựa trên các kernel trên:
  + **Linear**
* **Polynomial**
* **Radial basis**
* **Sigmoid**
* Sau khi train mô hình sử dụng tham số c từ 1 – 1.000.000.000. Ta được biểu đồ đánh từ sau c từ 1 đến 1.000.
* Từ biểu đồ trên ta nhận thấy được sự ổn định của kernel ‘linear’ khi áp dụng cho mô hình.
  + Sử dụng kernel linear với tham số c = 10 để tạo mô hình áp dụng cho bài toán.

1. **DEMO**

* Tạo giao diện web đơn giản bằng **streamlit**

Hình ảnh giao diện dự đoán cho :

*Gozila = [aquatic, toothed, backbone,breathes , fins, tail] + 2 legs*



Kết quả cho thấy Godzilla thuộc loài 3 - Bò sát.

**BẢNG PHÂN CÔNG**

|  |  |  |
| --- | --- | --- |
| Công việc | Thành viên | Ghi chú hỗ trợ |
| Thu thập dữ liệu | Cả nhóm |  |
| Phân tích dữ liệu | Đỗ Trung Nguyên | Các thành viên còn lại |
| Sử dụng libsvm | Cả nhóm |  |
| Đánh giá | Nguyễn Hưng | Các thành viên còn lại |
| Xây dựng giao diện ứng dụng | Cả nhóm |  |
| Slide thuyết trình | Đỗ Thành Công | Các thành viên còn lại |
| Báo cáo word | Cả nhóm |  |

**Github:** *https://github.com/tcong51/Advanced-Machine-Learning*