

Bài tập Lý thuyết 9

Thông tin sinh viên

Câu 1. Cho bộ dữ liệu như sau

Câu 2. Bảng dữ liệu bên dưới mô tả thống kê về thu nhập của sinh viên sau khi tốt nghiệp ở các trường. Trong đó “Thu nhập” là thuộc tính phân lớp.

Thông tin sinh viên

- MSSV: 22850034
- Họ và tên: Cao Hoài Việt
- Email: viet.ch2612@gmail.com

Câu 1. Cho bộ dữ liệu như sau

Name	Give Birth	Can Fly	Live in Water	Have Legs	Class
human	yes	no	no	yes	mammals
python	no	no	no	no	non-mammals
salmon	no	no	yes	no	non-mammals
whale	yes	no	yes	no	mammals
frog	no	no	sometimes	yes	non-mammals
komodo	no	no	no	yes	non-mammals
bat	yes	yes	no	yes	mammals
pigeon	no	yes	no	yes	non-mammals
cat	yes	no	no	yes	mammals
Leopard shark	yes	no	yes	no	non-mammals
turtle	no	no	sometimes	yes	non-mammals
penguin	no	no	sometimes	yes	non-mammals
porcupine	yes	no	no	yes	mammals
eel	no	no	yes	no	non-mammals
salamander	no	no	sometimes	yes	non-mammals
Gila monster	no	no	no	yes	non-mammals
platypus	no	no	no	yes	mammals
owl	no	yes	no	yes	non-mammals
dolphin	yes	no	yes	no	mammals
eagle	no	yes	no	yes	non-mammals

Dùng phương pháp Naïve Bayes (làm tròn laplace nếu cần), hãy cho biết kết quả phân lớp của các input sau

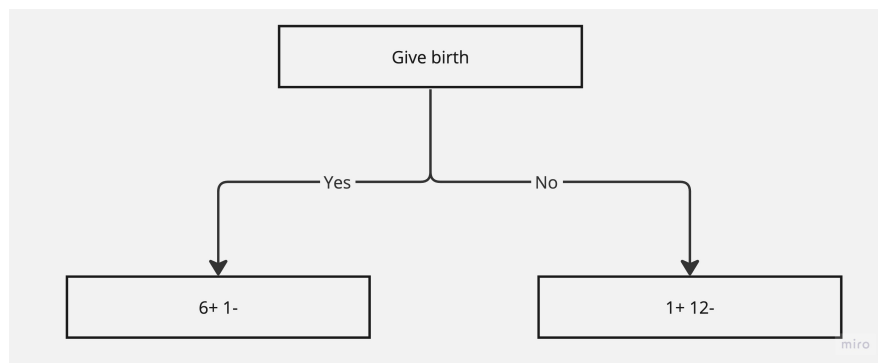
	Give Birth	Can Fly	Live in Water	Have Legs	Class
1	Yes	No	Yes	No	?
2	No	Yes	Sometimes	Yes	?
3	Yes	No	Yes	Yes	?
4	No	Yes	Yes	Yes	?

Vẽ cây ID3

Bước 1, tính Average Entropy

- **Give Birth = 0.461375**
- **Can Fly = 0.92578**
- **Live in Water = 0.78945**
- **Have Legs = 0.9337**

Ta sẽ chọn thuộc tính Give Birth vì có AE thấp nhất. Như vậy root sẽ là "Give Birth"



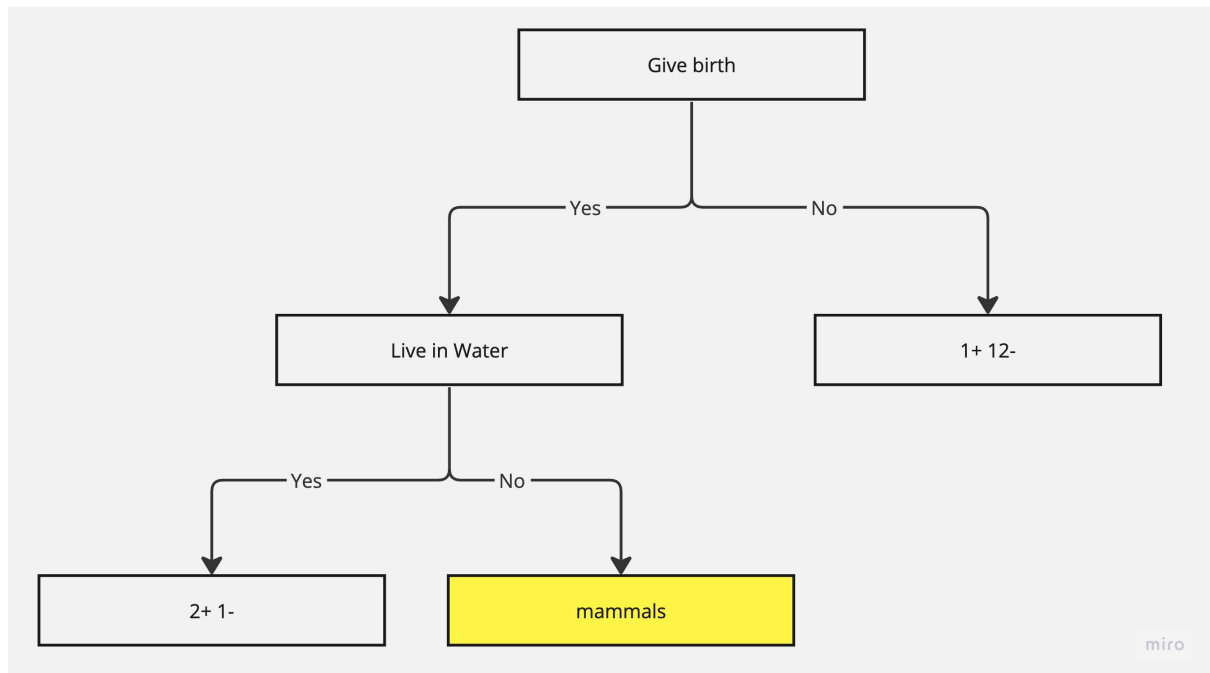
Ta tính nhánh Give Birth = Yes

Can Fly AE = 0.5571428571

Live in Water AE = 0.3935571429

Have Legs AE = 0.3935571429

Vì Live in Water = Have Legs nên chúng ta chọn đại Live in Water

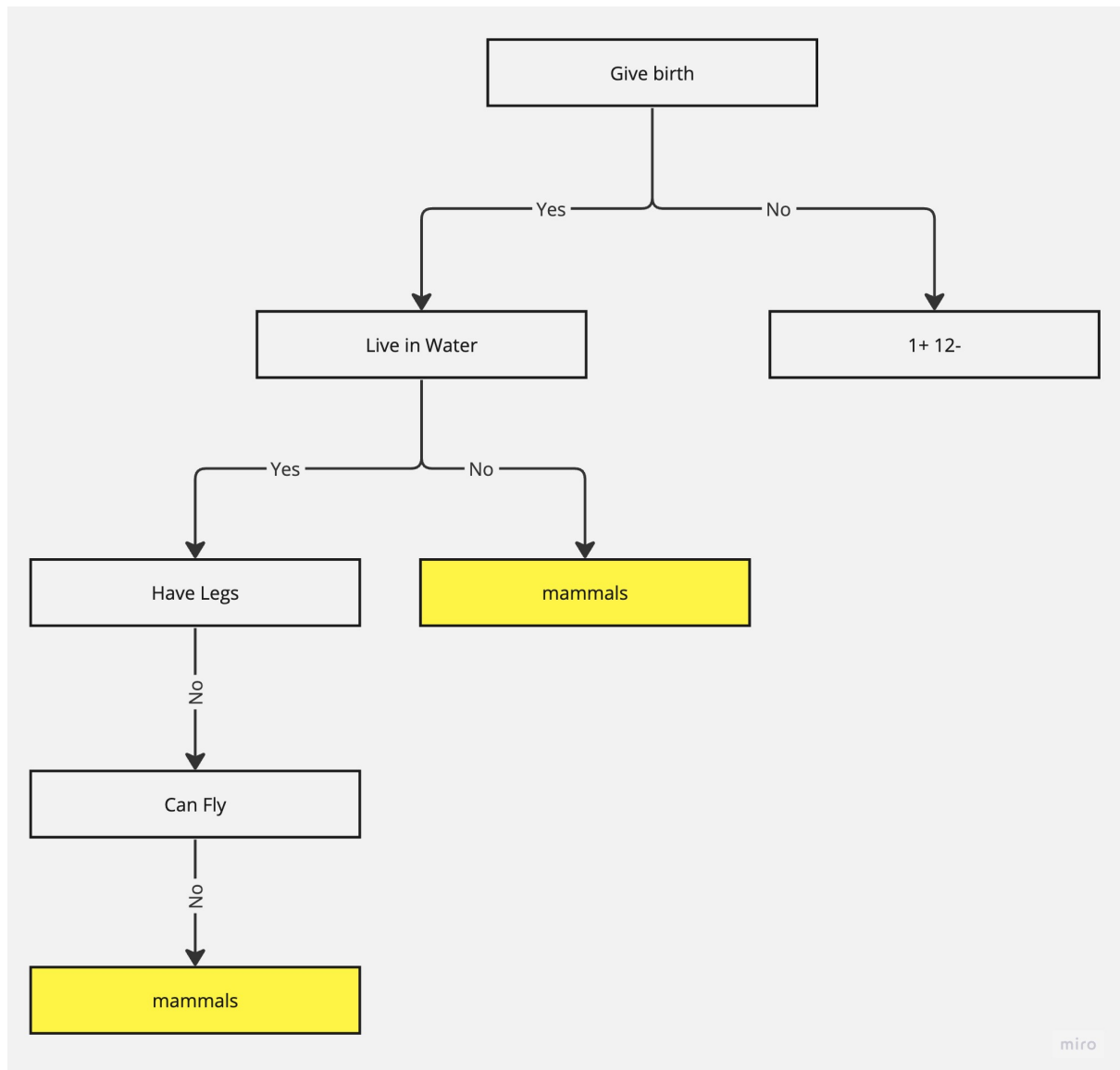


Nhánh Live in Water chưa rõ nên chúng ta sẽ xét tiếp

Can Fly AE= 0.9183

Have Legs AE= 0.9183

T sẽ chọn Have legs và Can Fly cuối cùng. Vì lá cuối cùng cũng chưa rõ (2+ 1-) nên ta sẽ lấy theo số đông, tức sẽ là **mammals**



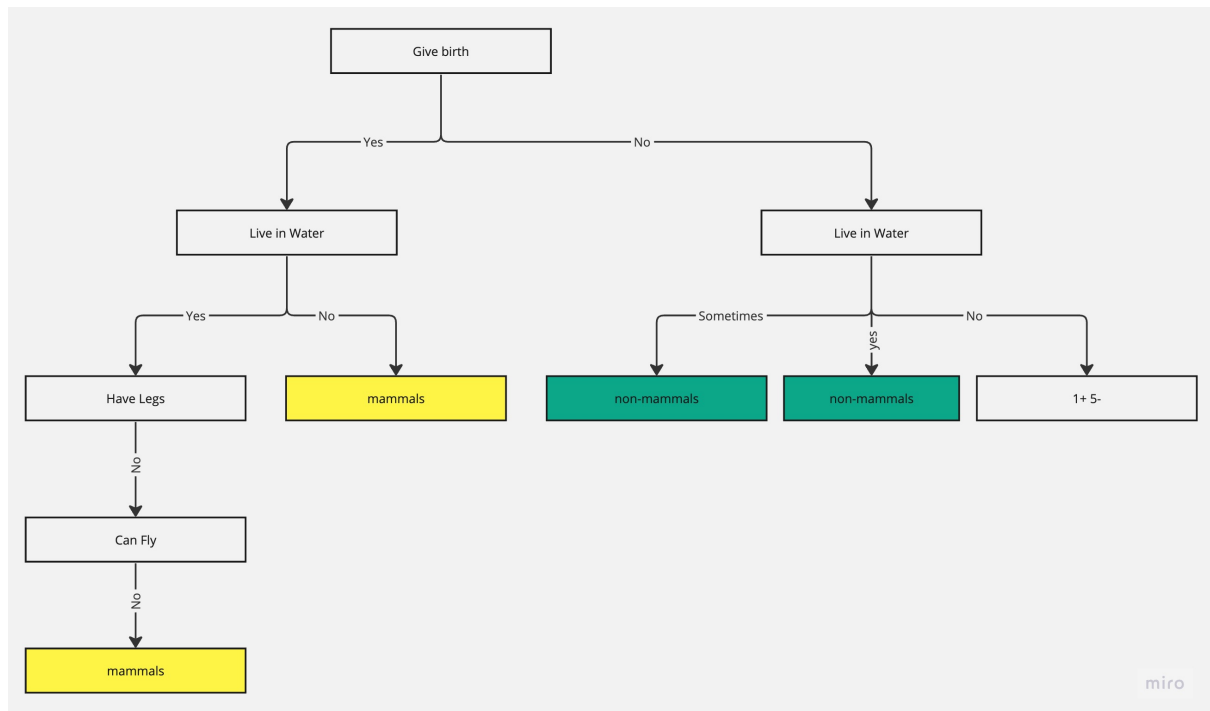
Tiếp theo đến nhánh Give Birth = NO

Can Fly AE = 0.3607692308

Live in Water AE = 0.35

Have Legs AE = 0.3607692308

Như vậy ta sẽ chọn tiếp là Live in Water

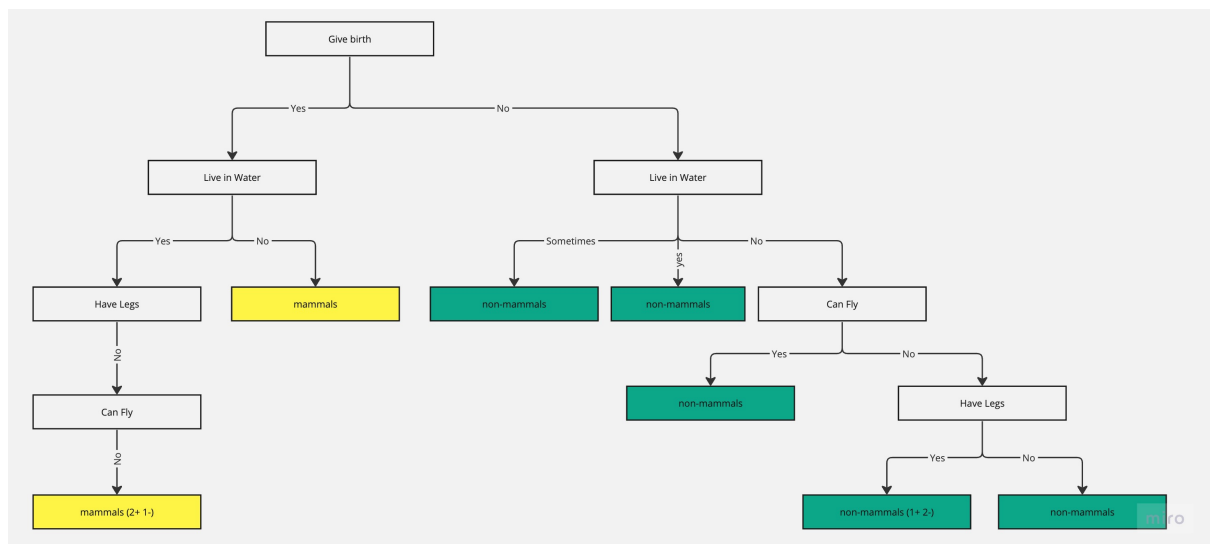


Ta tính tiếp thuộc tính Can Fly và Have Legs

Can Fly AE = 0.5247428571

Have Legs AE = 0.5571428571

Ta sẽ chọn theo Can Fly và cuối cùng là Have Legs



Câu 2. Bảng dữ liệu bên dưới mô tả thống kê về thu nhập của sinh viên sau khi tốt nghiệp ở các trường. Trong đó “*Thu nhập*” là thuộc tính phân lớp.

Mã SV	Trình độ học vấn	Nghề nghiệp	Số năm kinh nghiệm	Thu nhập
1712101	Cao đẳng	Lập trình	<3	Thấp
1412102	Cao đẳng	Lập trình	3...10	Thấp
1612103	Đại học	Lập trình	<3	Cao
1012104	Đại học	Quản lý dự án	>10	Thấp
1312105	Đại học	Quản lý dự án	3...10	Thấp
1412106	Đại học	Quản lý dự án	3...10	Cao
1012107	Đại học	Lập trình	>10	Cao
1512108	Đại học	Quản lý dự án	<3	Thấp
1012109	Cao đẳng	Lập trình	>10	Cao
1012110	Cao đẳng	Quản lý dự án	>10	Thấp

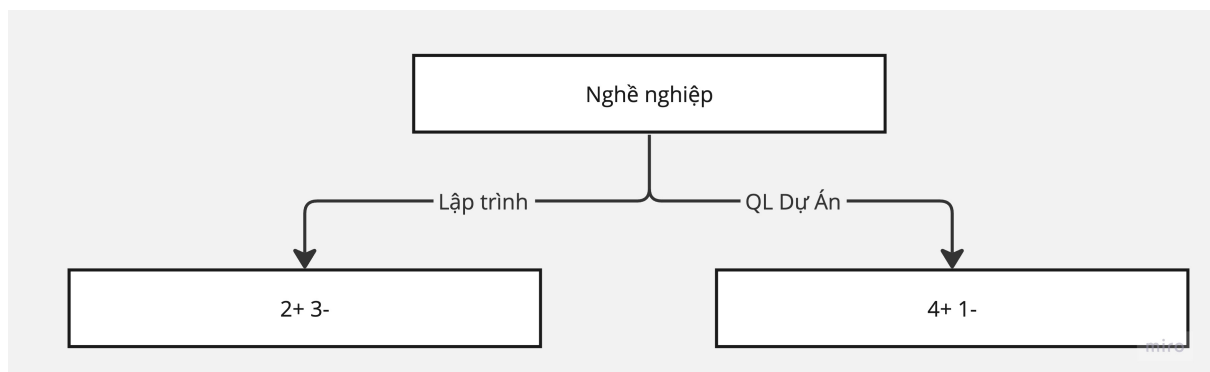
Ta tính Entropy của các thuộc tính với + là Thấp, - là Cao

Trình độ học vấn AE= **0.92452**

Nghề nghiệp AE= **0.8201**

Số năm kinh nghiệm AE= **0.95098**

Ta sẽ chọn gốc là Nghề nghiệp

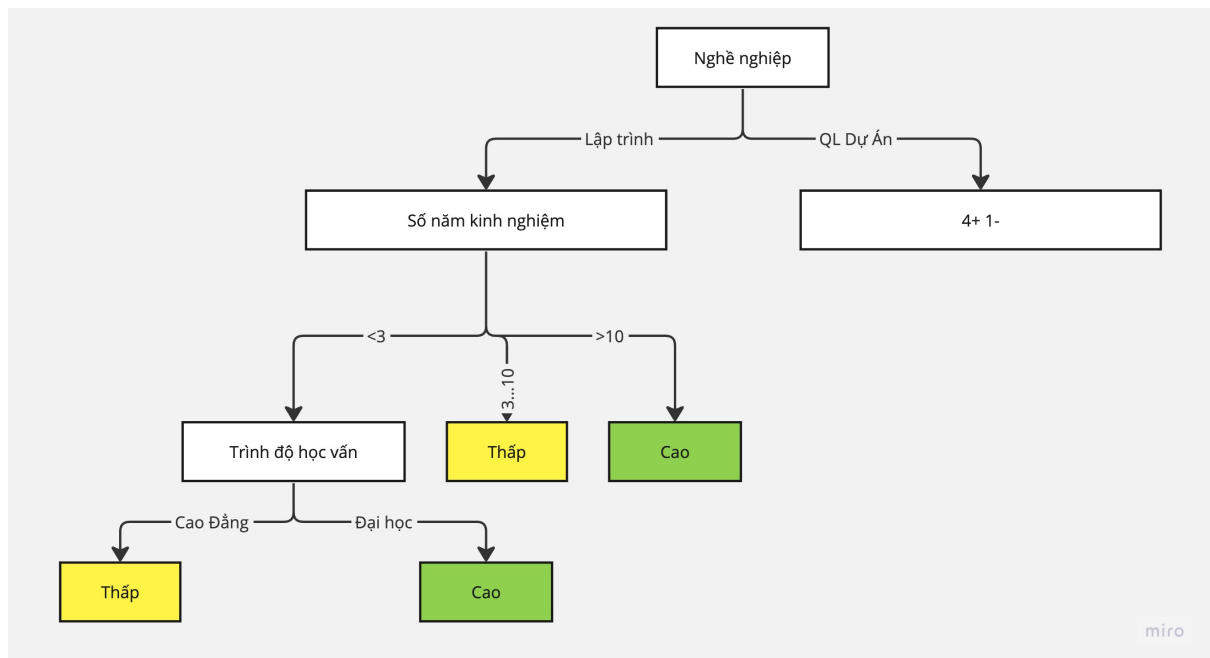
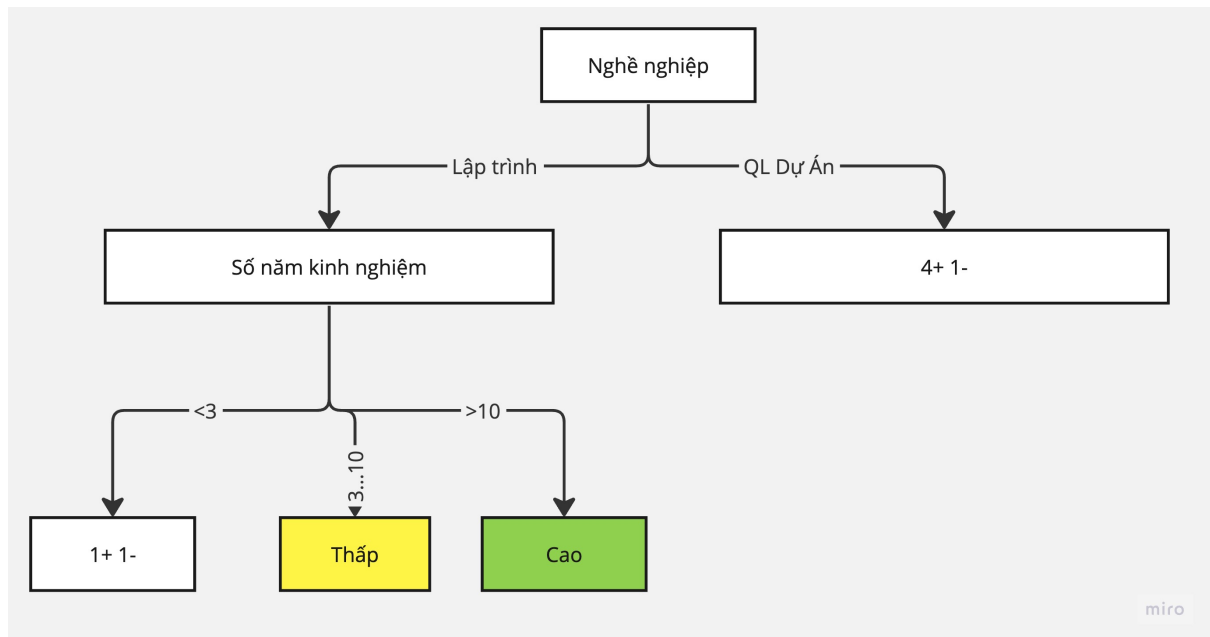


Ta tính tiếp trường hợp Nghề nghiệp là Lập trình

Trình độ học vấn AE= 0.55098

Số năm kinh nghiệm AE= **0.4**

Như vậy ta chọn thuộc tính tiếp theo là số năm kinh nghiệm và cuối cùng là Trình độ học vấn



Ta xét tới nhánh khi Nghề nghiệp = Quản lý Dự án

Trình độ học vấn AE= 0.64904

Số năm kinh nghiệm AE= 0.4

Ta chọn số năm kinh nghiệm rồi tới Trình độ học vấn

