

MSSV	Tên sinh viên
16520015	Nguyễn Xuân An
16521014	Võ Ngọc Quyển
17520178	Bùi Thị Thùy Trang
<b>17520505</b>	<b>Hà Đăng Hoàng</b>
18521471	Nguyễn Trọng Thuận

### Lab 02: Bài tập thực hành – câu 02

- Có 34781 loài cây.
- Có 69 vùng phân bố.
- Số loài cây trên mỗi vùng phân bố:

ab	3408	mn	3929	vt	3713
ak	2969	ms	4815	va	5638
ar	4610	mo	4638	vi	2185
az	6778	mt	4800	wa	5654
ca	11676	ne	3281	wv	4062
co	5465	nv	5670	wi	4321
ct	4391	nh	3635	wy	4710
de	3630	nj	4822	al	5702
dc	3080	nm	6403	bc	4875
fl	6621	ny	5773	mb	3023
ga	5942	nc	5926	nb	2856
hi	3804	nd	2682	lb	1433
id	5129	oh	4772	nf	2188
il	5167	ok	4651	nt	2024
in.	4440	or	7028	ns	2844
ia	3652	pa	181	nu	979
ks	3869	pr	4781	on	5068
ky	4555	ri	3295	pe	5515
la	5154	sc	5432	qc	4272
me	3969	sd	3185	sk	2846
md	5108	tn	4900	yt	2100

ma	4963	tx	8483	dengl	479
mi	4734	ut	6041	fraspm	1210

- d. Vùng có loài cây phân bố ít nhất:
- Tên: Pennsylvania (pa)
  - Số lượng: 181
  - Tỷ lệ: 33.57%
- e. Vùng có loài cây phân bố nhiều nhất:
- Tên: California (ca)
  - Số lượng: 11676
  - Tỷ lệ: 0.52%
- f. Trung bình một vùng có khoảng: 4347 cây

### Lab 02: Bài tập thực hành – câu 06

#### Large Itemsets L(5): ga,ms,al,nc,sc

1. [nc=y, al=y, sc=y, ms=y]: 3572 ==> [ga=y]: 3519 <conf:(0.99)> lift:(5.77)  
lev:(0.08) conv:(54.85)
2. [ga=y, nc=y, sc=y, ms=y]: 3612 ==> [al=y]: 3519 <conf:(0.97)> lift:(5.94)  
lev:(0.08) conv:(32.13)
3. [ga=y, nc=y, al=y, ms=y]: 3635 ==> [sc=y]: 3519 <conf:(0.97)> lift:(6.2)  
lev:(0.08) conv:(26.22)
4. [nc=y, sc=y, ms=y]: 3698 ==> [ga=y, al=y]: 3519 <conf:(0.95)> lift:(6.71)  
lev:(0.09) conv:(17.63)
5. [ga=y, al=y, sc=y, ms=y]: 3719 ==> [nc=y]: 3519 <conf:(0.95)> lift:(5.55)  
lev:(0.08) conv:(15.35)
6. [nc=y, al=y, ms=y]: 3733 ==> [ga=y, sc=y]: 3519 <conf:(0.94)> lift:(6.8)  
lev:(0.09) conv:(14.96)
7. [ga=y, nc=y, ms=y]: 3750 ==> [al=y, sc=y]: 3519 <conf:(0.94)> lift:(7.26)  
lev:(0.09) conv:(14.07)
8. [al=y, sc=y, ms=y]: 3805 ==> [ga=y, nc=y]: 3519 <conf:(0.92)> lift:(6.66)  
lev:(0.09) conv:(11.42)

#### Large Itemsets L(5): ga,va,al,nc,sc

1. [ga=y, al=y, va=y, sc=y]: 3579 ==> [nc=y]: 3529 <conf:(0.99)> lift:(5.79)  
lev:(0.08) conv:(58.22)
2. [nc=y, al=y, va=y, sc=y]: 3608 ==> [ga=y]: 3529 <conf:(0.98)> lift:(5.73)  
lev:(0.08) conv:(37.4)
3. [al=y, va=y, sc=y]: 3682 ==> [ga=y, nc=y]: 3529 <conf:(0.96)> lift:(6.9)  
lev:(0.09) conv:(20.59)
4. [ga=y, nc=y, al=y, va=y]: 3694 ==> [sc=y]: 3529 <conf:(0.96)> lift:(6.12)  
lev:(0.08) conv:(18.78)
5. [ga=y, al=y, va=y]: 3788 ==> [nc=y, sc=y]: 3529 <conf:(0.93)> lift:(6.75)  
lev:(0.09) conv:(12.56)
6. [ga=y, nc=y, va=y, sc=y]: 3825 ==> [al=y]: 3529 <conf:(0.92)> lift:(5.63)  
lev:(0.08) conv:(10.77)

Việc đánh giá các luật kết hợp trên đặt ra hai mối quan tâm lớn. Đầu tiên là tính Scalable (tạm dịch: khả năng mở rộng), trong đó đề cập đến khả năng của một thuật toán để xử lý một lượng lớn dữ liệu theo cách hiệu quả phù hợp. Tiếp đến là khả năng ứng dụng, chẳng hạn như ý nghĩa sinh học, ý nghĩa đối với thực tiễn xã hội.

- Đối với khả năng mở rộng là một vấn đề quan trọng khác trong nghiên cứu y sinh nói riêng và thực tiễn nói chung. Thông thường, kết hợp kiến thức sinh học đã biết trước đây với các thuật toán khai thác quy tắc kết hợp được coi là cung cấp kết quả có ý nghĩa sinh học tốt hơn; tuy nhiên, trong bài thực hành này, thuật toán sẽ có xu hướng phù hợp với kiến thức sinh học hơn nên có thể làm giảm khả năng có được các quy tắc chưa được khám phá.
- Khả năng ứng dụng được rút ra từ các luật kết hợp trên có ý nghĩa đối với thực tiễn:
  - Từ các luật thu về, ta có thể phán đoán được nơi sống và môi trường sống phù hợp cho các loài thực vật.  
⇒ Từ đó dễ dàng có biện pháp khai thác, bảo vệ và phát triển.
  - Các luật có xuất hiện 1 vùng với xác suất cao thì có thể kết luận đây là kiểu môi trường phù hợp với đa số các loại cây hiện tại.  
⇒ Từ đó có cơ sở phát triển và xây dựng môi trường tự nhiên vùng này.

