Bài 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 trơn 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 | ? |

**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 |

Hãy áp dụng phương pháp Naïve Bayes (sửa lỗi Laplace nếu cần) để xác định lớp cho hai mẫu dữ liệu sau:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Mã SV** | **Trình độ học**  **vấn** | **Nghề nghiệp** | **Số năm kinh**  **nghiệm** | **Thu nhập** |
| 1512111 | Đại học | Lập trình | 3…10 | ? |
| 1712112 | Cao đẳng | Quản lý dự án | <3 | ? |