**BÀI TẬP 4**

**Bài 1.** Có 15 học sinh có điểm thi ba môn Toán Lý, Hóa và kết quả Nguyện vọng Đại học như sau:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **MaHS** | **Toan** | **Ly** | **Hoa** | **ĐH** |
| 1 | Giỏi | Giỏi | Giỏi | Đỗ |
| 2 | Khá | TB | TB | Trượt |
| 3 | Giỏi | Giỏi | Khá | Đỗ |
| 4 | TB | Giỏi | Giỏi | Đỗ |
| 5 | Giỏi | Giỏi | TB | Đỗ |
| 6 | TB | Khá | Khá | Đỗ |
| 7 | Khá | Giỏi | Giỏi | Đỗ |
| 8 | TB | Khá | TB | Trượt |
| 9 | Khá | TB | Giỏi | Đỗ |
| 10 | Giỏi | TB | TB | Trượt |
| 11 | TB | TB | Giỏi | Trượt |
| 12 | Giỏi | Khá | TB | Đỗ |
| 13 | Khá | Giỏi | TB | Đỗ |
| 14 | Giỏi | TB | Khá | Đỗ |
| 15 | TB | TB | TB | Trượt |

1. Yêu cầu thực hiện bằng tay:
2. Xây dựng cây quyết định bằng thuật toán ID3 sử dụng độ đo Entropy
3. Xây dựng cây CART sử dụng chỉ số Gini Index
4. Thực hiện lập trình với thư viện sklearn với phương pháp cây quyết định và rừng ngẫu nhiên để so sánh kết quả

Bài 2. Điều tra 12 người về độ tuổi, tình trạng hôn nhân, bất động sản, thu nhập và đánh giá mức độ rủi ro tài chính được như sau:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ID | Tuổi | Hôn nhân | BDS | Thu nhập | Rủi ro tài chính |
| 1 | Trẻ | Độc thân | Nhà thuê | Thấp | Không |
| 2 | Trung bình | Đã kết hôn | Nhà sở hữu | Cao | Không |
| 3 | Trẻ | Đã kết hôn | Nhà thuê | Thấp | Có |
| 4 | Trung bình | Độc thân | Nhà thuê | Thấp | Có |
| 5 | Trung bình | Đã kết hôn | Nhà sở hữu | Thấp | Có |
| 6 | Trẻ | Đã kết hôn | Nhà sở hữu | Thấp | Có |
| 7 | Trẻ | Độc thân | Nhà thuê | Trung bình | Không |
| 8 | Trung bình | Đã kết hôn | Nhà sở hữu | Cao | Không |
| 9 | Cao | Độc thân | Nhà sở hữu | Trung bình | Có |
| 10 | Cao | Đã kết hôn | Nhà sở hữu | Trung bình | Không |
| 11 | Trẻ | Độc thân | Nhà thuê | Thấp | Có |
| 12 | Trung bình | Đã kết hôn | Nhà sở hữu | Thấp | Có |

1. Yêu cầu thực hiện bằng tay:
2. Xây dựng cây quyết định bằng thuật toán ID3 sử dụng độ đo Entropy
3. Xây dựng cây CART sử dụng chỉ số Gini Index
4. Thực hiện lập trình với thư viện sklearn với phương pháp cây quyết định và rừng ngẫu nhiên để so sánh kết quả

Bài 3. Điều tra 15 người về độ tuổi, tình trạng hôn nhân, bất động sản, thu nhập và đánh giá mức độ rủi ro tín dụng được như sau:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ID | Độ tuổi | Hôn nhân | Sở hữu BĐS | Thu nhập | Rủi ro tín dụng |
| 1 | 25 | Độc thân | Ở cùng bố mẹ | 7 | 0 |
| 2 | 40 | Đã kết hôn | Nhà sở hữu | 18 | 0 |
| 3 | 35 | Từng ly hôn | Nhà thuê | 12 | 1 |
| 4 | 27 | Đã kết hôn | Ở cùng bố mẹ | 9 | 1 |
| 5 | 31 | Độc thân | Nhà thuê | 6 | 1 |
| 6 | 36 | Đã kết hôn | Nhà sở hữu | 8 | 1 |
| 7 | 48 | Độc thân | Nhà thuê | 7 | 0 |
| 8 | 26 | Đã kết hôn | Nhà thuê | 8 | 1 |
| 9 | 33 | Từng ly hôn | Ở cùng bố mẹ | 5 | 1 |
| 10 | 29 | Độc thân | Nhà thuê | 10 | 0 |
| 11 | 38 | Đã kết hôn | Nhà sở hữu | 16 | 0 |
| 12 | 44 | Độc thân | Nhà sở hữu | 14 | 1 |
| 13 | 42 | Đã kết hôn | Nhà sở hữu | 10 | 0 |
| 14 | 28 | Độc thân | Nhà thuê | 7 | 1 |
| 15 | 30 | Đã kết hôn | Ở cùng bố mẹ | 6 | 1 |

1. Yêu cầu thực hiện bằng tay:
2. Xây dựng cây quyết định bằng thuật toán ID3 sử dụng độ đo Entropy
3. Xây dựng cây CART sử dụng chỉ số Gini Index
4. Thực hiện lập trình với thư viện sklearn với phương pháp cây quyết định và rừng ngẫu nhiên để so sánh kết quả

Gợi ý: Tiến hành chuyển dữ liệu độ tuổi và thu nhập sang kiểu định danh (rời rạc) theo K nhóm (ví dụ K =3)