## Yêu cầu 2:

* Sử dụng thuật toán J48 để huấn luyện tập dữ liệu.
  + Các tham số cần thiết để chạy thuật toán:

|  |  |
| --- | --- |
| confidenceFactor | 0.25 |
| minNumObj | 2 |
| numFolds | 3 |
| seed | 1 |

* + Thuật toán này không gặp bất cứ vấn đề nào khi chạy.
  + Cây quyết định của thuật toán:



* + Đánh giá hiệu quả phân lớp của thuật toán đối với tập dữ liệu được cho theo hai phương pháp:
    - Cross-validation: tỉ lệ phân lớp đúng luôn là 100% và số mẫu phân lớp đúng chính là số mẫu ban đầu của tập dữ liệu với mọi giá trị của Folds.

|  |  |  |
| --- | --- | --- |
|  | Số mẫu | Tỉ lệ (%) |
| Phân lớp đúng | 8124 | 100 |
| Phân lớp sai | 0 | 0 |
| Không phân lớp được | 0 | 0 |
| Tổng | 8124 | 100 |

* + - Percentage split:khi cho tỉ lệ phân chia thấp thì có tình trạng phân lớp sai. Tỉ lệ phân chia là 35% thì đạt hiệu quả phân lớp cao nhất.
      * Một số kết quả thử nghiệm:

|  |  |
| --- | --- |
| Percentage split |  |
| 1% | |  |  |  | | --- | --- | --- | |  | Số mẫu | Tỉ lệ (%) | | Phân lớp đúng | 7924 | 98,5205 | | Phân lớp sai | 119 | 1,4795 | | Không phân lớp được | 0 | 0 | | Tổng | 8043 | 100 | |
| 25% | |  |  |  | | --- | --- | --- | |  | Số mẫu | Tỉ lệ (%) | | Phân lớp đúng | 6086 | 99,8851 | | Phân lớp sai | 7 | 0,1149 | | Không phân lớp được | 0 | 0 | | Tổng | 6093 | 100 | |
| 34% | |  |  |  | | --- | --- | --- | |  | Số mẫu | Tỉ lệ (%) | | Phân lớp đúng | 5355 | 99,8695 | | Phân lớp sai | 7 | 0,1305 | | Không phân lớp được | 0 | 0 | | Tổng | 5362 | 100 | |
| 35% | |  |  |  | | --- | --- | --- | |  | Số mẫu | Tỉ lệ (%) | | Phân lớp đúng | 5281 | 100 | | Phân lớp sai | 0 | 0 | | Không phân lớp được | 0 | 0 | | Tổng | 5281 | 100 | |
| 40% | |  |  |  | | --- | --- | --- | |  | Số mẫu | Tỉ lệ (%) | | Phân lớp đúng | 4875 | 100 | | Phân lớp sai | 0 | 0 | | Không phân lớp được | 0 | 0 | | Tổng | 4875 | 100 | |
| 50% | |  |  |  | | --- | --- | --- | |  | Số mẫu | Tỉ lệ (%) | | Phân lớp đúng | 4062 | 100 | | Phân lớp sai | 0 | 0 | | Không phân lớp được | 0 | 0 | | Tổng | 4062 | 100 | |