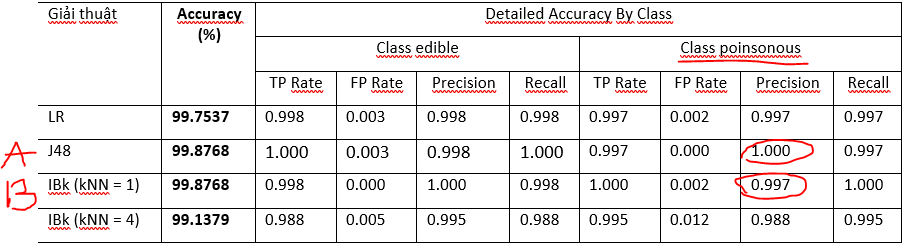
1. Lập bảng thống kê độ chính xác phân lớp của mỗi giải thuật trên dữ liệu được cho.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Giải thuật | **Accuracy**  **(%)** | Detailed Accuracy By Class | | | | | | | |
| Class edible | | | | Class poinsonous | | | |
| TP Rate | FP Rate | Precision | Recall | TP Rate | FP Rate | Precision | Recall |
| LR | **99.7537** | 0.998 | 0.003 | 0.998 | 0.998 | 0.997 | 0.002 | 0.997 | 0.997 |
| J48 | **99.8768** | 1.000 | 0.003 | 0.998 | 1.000 | 0.997 | 0.000 | 1.000 | 0.997 |
| IBk (kNN = 1) | **99.8768** | 0.998 | 0.000 | 1.000 | 0.998 | 1.000 | 0.002 | 0.997 | 1.000 |
| IBk (kNN = 4) | **99.1379** | 0.988 | 0.005 | 0.995 | 0.988 | 0.995 | 0.012 | 0.988 | 0.995 |

1. Giả sử Weka cho báo cáo hiệu quả của 2 giải thuật A và B

|  |  |  |
| --- | --- | --- |
| Giải thuật A | | |
| Accuracy: 99.8768% | | |
| Class | TP Rate | FP Rate |
| edible | 1,000 | 0,003 |
| poisonous | 0,997 | 0,000 |

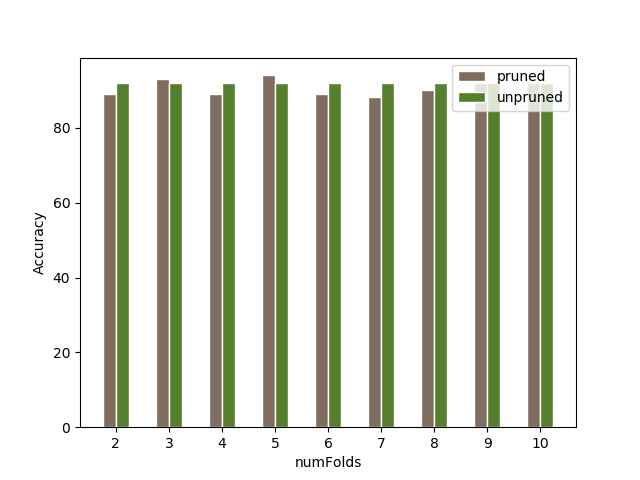
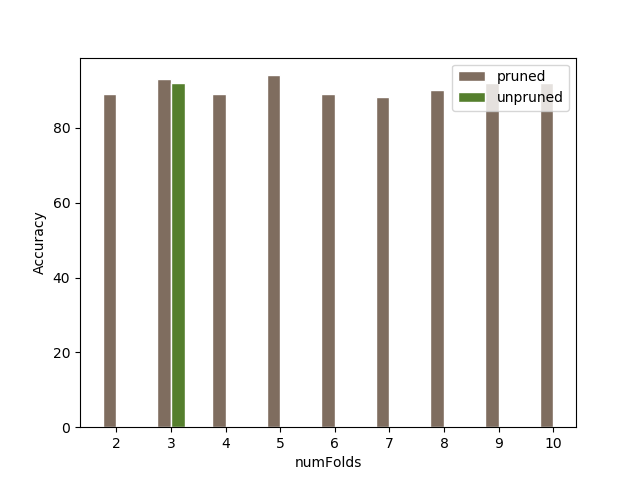
|  |  |  |
| --- | --- | --- |
| Giải thuật B | | |
| Accuracy: 99.8768% | | |
| Class | TP Rate | FP Rate |
| edible | 0,998 | 0 |
| poisonous | 1 | 0,002 |



Có thể tính precision cho poisonous, nhận xét thấy giải thuật A cho precision cao hơn so với giải thuật B. Như trên hình giải thuật A tương ứng với J48, giải thuật B tương ứng với IBk (kNN = 1).

1. Vẽ đồ thị thể hiện độ chính xác phân lớp (trục tung) theo sự biến thiên của numFolds (trục hoành), xét xả 2 trường hợp có tỉa nhánh và không tỉa nhánh (sử dụng python).

Chạy file 3.py trong phần src để vẽ đồ thị



(3 và 3\_1)

2. Tham số numFolds có vai trò trong J48:
3. Mô tả sự tác động của tham số này đến hiệu quả của cây quyết định xây dựng được khi thay đổi giá trị tham số
5. Dựa vào đồ thị, bình luận về hiệu quả tỉa nhánh giảm lỗi trên cây quyết định thu được
6. Tỉa nhánh cây giúp cải thiện độ chính xác (tức là hiện tượng máy học nào đang xảy ra)
7. Đánh giá độ chính xác của mô hình được chọn

Chọn SVM (weka.classifiers.functions.SMO)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Class | Detailed Accuracy | | | |
| TP Rate | FP Rate | Precision | Recall |
| A | 0.998 | 0.000 | 0.994 | 0.998 |
| B | 0.967 | 0.002 | 0.945 | 0.967 |
| C | 0.990 | 0.001 | 0.983 | 0.990 |
| D | 0.964 | 0.002 | 0.942 | 0.964 |
| E | 0.976 | 0.001 | 0.966 | 0.976 |
| F | 0.968 | 0.001 | 0.972 | 0.968 |
| G | 0.966 | 0.001 | 0.974 | 0.966 |
| H | 0.933 | 0.002 | 0.948 | 0.933 |
| I | 0.970 | 0.001 | 0.970 | 0.970 |
| J | 0.958 | 0.001 | 0.972 | 0.958 |
| K | 0.972 | 0.002 | 0.961 | 0.972 |
| L | 0.977 | 0.000 | 0.988 | 0.977 |
| M | 0.983 | 0.001 | 0.986 | 0.983 |
| N | 0.980 | 0.001 | 0.978 | 0.980 |
| O | 0.969 | 0.001 | 0.978 | 0.969 |
| P | 0.965 | 0.001 | 0.976 | 0.965 |
| Q | 0.986 | 0.001 | 0.983 | 0.986 |
| R | 0.960 | 0.001 | 0.962 | 0.960 |
| S | 0.992 | 0.000 | 0.997 | 0.992 |
| T | 0.983 | 0.000 | 0.994 | 0.983 |
| U | 0.992 | 0.001 | 0.988 | 0.992 |
| V | 0.977 | 0.001 | 0.977 | 0.977 |
| W | 0.990 | 0.000 | 0.992 | 0.990 |
| X | 0.987 | 0.001 | 0.982 | 0.987 |
| Y | 0.992 | 0.001 | 0.988 | 0.992 |
| Z | 0.991 | 0.000 | 0.993 | 0.991 |

1. Mô tả ngắn gọn về mặt lý thuyết giải thuật được chọn để xây dựng mô hình. Vì sao cho rằng đây là giải thuật tốt nhất so với các giải thuật khác?

* Giải thuật

1. Báo cáo với bộ tham số thực nghiệm. Cấu hình tham số ảnh hưởng như thế nào đến hiệu quả của giải thuật?