분석에는 UCI에서 제공하는 ‘Mushrooms’ 라는 독버섯 판별 데이터를 이용하였다.

* 데이터 출처 : UC Lrvine Machine Learning Repository (https://archive.ics.uci.edu/ml/datasets/Mushroom)
* 북 아메리카에서 자란 버섯의 식용여부와 총 22가지의 버섯 특징 데이터
* 8124 개의 버섯 데이터

**변수 이름과 설명**

|  |  |
| --- | --- |
| 1. cap-shape : 갓 형태 2. cap-surface : 갓의 표면 3. cap-color : 갓의 색깔 4. bruises : 멍 유무 5. odor : 냄새 6. gill-attachment : 주름살 7. gill-spacing : 주름살 간격 8. gill-size : 주름살 크기 9. gill-color : 주름살 색 10. stalk-shape : 줄기 모양 11. stalk-root : 줄기 뿌리 | 1. stalk-surface-above-ring : 턱받이 위쪽 줄기 모양 2. stalk-surface-below-ring : 턱받이 아래쪽 줄기 모양 3. stalk-color-above-ring : 갓 위쪽의 줄기 색깔 4. stalk-color-below-ring : 갓 아래쪽의 줄기 색깔 5. veil-type : 베일 유형 6. veil-color : 베일 색깔 7. ring-number : 턱받이 갯수 8. ring-type : 턱받이 유형 9. spore-print-color : 포자 색깔 10. population : 서식 분포 유형 11. habitat : 서식지 |

주름살 (gill) : 생식세포인 홀씨를 생성하는 버섯의 한 부분. 갓의 아랫면에 있다.

대주머니 / 균포 (volva) : 미성숙한 버섯을 완전히 감싸고 있던 막이 줄기가 생장함에 따라 찢어진 흔적.

균사 / 팡이실 (hypha) : 간혹 흰색을 띠는 미세한 사상체. 버섯의 생장에 필요한 물과 유기물질을 흡수한다.

균사체 (mycelium) : 홀씨의 발아를 통해 생성된 균사가 서로 얽힌 집합체. 여기에서 버섯의 지상 부분이 발달한다.

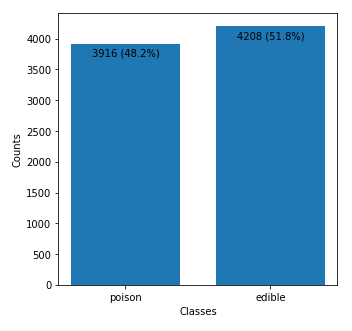
홀씨 / 포자 (spores) : 일반적으로 대기로 방출된 뒤 기질에 떨어져서 새로운 버섯이 생겨나게 하는 미세한 생식세포.

줄기 (stem) : 버섯의 갓을 지지하는 기둥.

턱받이 (ring) : 갓의 아래에서 줄기를 둘러싼 막. 미성숙한 버섯의 주름살을 감싸고 있다가 갓이 성장하면서 파열된 막의 흔적이다.

갓 (cap) : 형태와 색상이 다양한 버섯의 상단부. 주름살을 보호한다. 대개 머리에 쓰는 모자를 닮은 형태를 띠는 데서 그 이름이 비롯되었다.

**버섯 데이터가 한 쪽으로 편향 되어 있을까?**

해당 버섯 판별 문제는 binary classification 문제기 때문에 당연하게 클래스가 e, edible 이면, 먹을 수 있는 버섯이고, p, poison 이면, 먹을 수 없는 독버섯 이라고 할 수 있다. 데이터를 사용하기 이전에 해당 클래스마다 몇 개의 데이터가 있는지 한 쪽으로 몰려 있는지 알아보고 만약 편향이 되어 있으면 데이터읙 균형을 맞춰주는 작업을 해야 한다. 아래 데이터 분포를 보면 버섯 데이터 균형은 edible 이 51.8%, poison 이 48.2% 로 데이터 분포가 매우 균등하다고 볼 수 있다.

|  |  |  |
| --- | --- | --- |
| **Class** | **Poison** | **Edible** |
| **버섯 개수** | 3916 | 4208 |
| **(%)** | 48.2% | 51.8% |

**독버섯과 식용 가능 버섯의 특징은 서로 다를까?**