**Résultats du test technique et voies d’amélioration :**

J’ai écrit deux programmes pour répondre à la problématique posée :

* Tout d’abord, un premier programme d’entraînement qui détermine la dangerosité de chaque feature. Celui-ci fonctionne en deux parties : dans un premier temps, une boucle balaye le dataset et range les features du champignon dans une liste « pois » ou dans une liste « com ». Dans un second temps, une autre boucle attribue un taux de dangerosité entre 0 et 1 à chaque feature en comptant le nombre de fois que celle-ci apparaît sur un champignon empoisonné. Ce programme retourne une liste de tuples sous la forme « (feature, taux de dangerosité) ». Cependant, il n’est malheureusement pas bien optimisé car il consomme beaucoup de ressources processeur et met également beaucoup de temps à fonctionner (il vaut mieux éviter de l’exécuter car je n’ai jamais pu en obtenir un quelconque résultat à cause de son temps d’exécution).
* Ensuite, un deuxième programme qui permet, à partir d’un dataset et de la liste renvoyée par le programme d’entraînement, de déterminer la dangerosité d’un champignon grâce au taux de dangerosité de ses features. Si un champignon a un taux total de dangerosité de 0,5 ou plus, il est déterminé empoisonné (il vaut mieux prédire qu’un champignon qui a une chance sur deux de ne pas être comestible ne l’est pas afin de ne pas prendre de risque). Sinon, il est déterminé comme étant comestible. De plus, si une de ses features est de manière sûre caractéristique d’un champignon empoisonné (taux de 1) ou d’un champignon comestible (taux de 0), alors le champignon est directement déterminé comme étant respectivement empoisonné ou comestible. Ce programme ne fonctionne pas car je n’ai pas trouvé de moyen d’obtenir les taux des features (lignes 42, 45, 49) à partir d’une liste de tuples de la forme « (feature, taux) ».

Cette méthode a pour avantage d’être précise car elle prend en compte toutes les caractéristiques du champignon. Cependant, elle est lourde à utiliser et ne fonctionne pas en l’état, il faudrait donc sélectionner uniquement les caractéristiques les plus importantes afin d’alléger son exécution et trouver des moyens de l’optimiser davantage. Il peut également être intéressant d’identifier des combinaisons de caractéristiques apparaissant souvent sur un champignon empoisonné ou sur un champignon comestible.