Project : Data science and IA Python

Emilie CHEVREAU, Valentin Ré and François Romano

WHY?

It can be challenging to determine whether a mushroom is edible or toxic without specialized knowledge. Indeed, it is crucial to assess the toxicity of mushrooms before consuming them. Currently, there is a list of 140,000 documented mushroom species, making it difficult to recognize each variety. Distinguishing between toxic and non-toxic mushrooms can be a risky task, so it is the responsibility of our project to address this.

WHAT?

How can one determine the toxicity or edibility of a mushroom based on a simple photo of its physical characteristics?

There are already applications that, through an image, can indicate whether a mushroom is toxic or edible. However, our goal is not to offer a binary method. We aim to help users visually distinguish the differences and similarities between an edible mushroom and a toxic one (smell and touch can be important but challenging factors to process). This approach will focus on educating users about specific mushroom characteristics. The application can provide a toxicity percentage for the collected mushroom. The application will also provide insights into its prognosis, meaning it will offer its expertise and reasons for why it believes the mushroom is toxic or not. This will enable users to become more proficient in this field through the analyses provided by our application.

HOW?

We would like to create an offline application with a pre-trained AI to determine the toxicity percentage of a mushroom. We will use an existing mushroom database that contains visual characteristics.

The app will not be designed to identify the species of the mushroom but rather its visual characteristics, including color, stem, cap, shape, spots, size, and odor. First of all, with the database, we need to identify the common features of mushrooms.

Methods:

PCA (Principal Component Analysis) to identify commonalities among toxic and edible mushrooms.

EM (Expectation-Maximization) algorithm : it may be used to estimate model parameters for mushroom toxicity classification

MLE (Maximum Likelihood Estimation).

DIA (Discriminant Analysis) : distinguishing between different classes of data, in this case, toxic and non-toxic mushrooms

CNN (Convolutional Neural Network) : lass of deep learning models particularly well-suited for image classification

VGG (Visual Geometry Group).

RNN (Recurrent Neural Network).

YOLO (You Only Look Once) object detection method : is an object detection technique used to identify and locate objects within images. It can be valuable for recognizing specific mushroom features

Image segmentation techniques : isolating specific mushroom characteristics

GAN (Generative Adversarial Network).

SVM (Support Vector Machine) : structured data and can be beneficial for distinguishing between toxic and non-toxic mushrooms

Dataset and Reaserch exist :

<https://www.kaggle.com/datasets/uciml/mushroom-classification>

<https://www.kaggle.com/code/auxeno/mushroom-hunting-a-simplified-guide>

https://www.kaggle.com/code/ahmedtronic/mashroom-classification

It includes a CSV file containing various features of mushrooms

Read images

And extract features from images

Match the data : classification and the species

Species -> features