1. Introduce it to us (where do you find it? Who compiled it? number of records, attributes ...).
We found this dataset on Kaggle. It is a website that has a lot of datasets of various categories. Kaggle supports science projects, and you can build the code and data in this model to help users doing machine learning. And on this website we were interested in the mushroom data, so we got it as our project research.

The research is from the Audubon Society Field Guide to North American Mushrooms. This dataset is all about the appearance of the mushrooms. It recorded the shape of the cap, the cap's color, the size of the mushroom…… and so on.
Number of instances: 8124
Number of attributes: 22
Dataset : Mushroom

2. what kind of machine learning project we can do with this dataset.
We know that there are over thousands of mushrooms in the world, and it's almost impossible to identify them all as a non-professional. We wonder if there is any pattern to learn what kind of mushroom is non-toxic. So we would like to use this dataset to recognize if this mushroom is edible.

Then, find and introduce a machine learning software (other than WEKA) or a machine learning library to us.

ML software: **KNIME**

KNIME Analytics Platform is an open source data science software. It integrates a large number of the latest developments in data science and artificial intelligence technologies and transforms them into one boxed component after another. Using KNIME to read data, you can use it to generate tables and graphs or to predict unknown values based on pre-built models.

Roughly speaking, it can be divided into the first half of the establishment of the collection and cleaning of data, modeling and integration, and the second half of the practical application of deployment and management, understanding and optimization.

One of the features of KNIME is the integration of a large number of machine learning and artificial intelligence techniques. For building machine learning models, classification, regression, dimensionality reduction, binning, deep learning, tree-based model algorithms, and logistic regression can be used.