Buisness understanding

Mushroom hunting is a hobby popular in Estonia, Russia, Poland and many other countries. It implies going a far distance from cities to some forest and gathering eadible mushrooms (to eat them afterwards). It is common to spend 3-5 or more hours in the forest and to gather about 10 liters of mushrooms, altrough the volume depend greatly on how lucky and trained the hunter is. Of course, not all mushrooms that can be found in the forest are eadible, and this is a major inconvinience, since an inexperienced gatherer could pick all the mushrooms and later (after consulting with experienced gatherers) find out that some or even most of them are not eadible.

Given the small dataset we would not try to build a system that will be sufficient to determine if a mushroom is safe to eat, since wrongly assuming that a mushroom is safe could end up badly (need for hospitalization). This may only change if we find a larger dataset that will be homogeneous-enough with ours. Instead, we would like to build a system (probably an application) that will allow an inexperienced mushroom hunter to gain some experience and to learn how ineadible mushrooms look, so that he/she will not pick them up. This means that afterward consulting is still necessary. Given that, in the long run the system could ask the hunters to provide their data and therefore become better; possibly one day eliminating the need of consulting. The system should determine if a mushroom is eadible in under a second on a personal computer, learning time is not bounded strictly (a few hours on a personal computer). Usability questions (for example: "Will there be an Android app for this and will it be easy to use to beginner hunters?") are currently outside of the scope of the project.

Our project success criteria: we may make a pool for inexperienced shroomers who used our system and see if our system made a difference for them. If it would - great!

In this project we will use Kaggle dataset (https://www.kaggle.com/uciml/mushroom-classification) and our personal computers. In a long run, we could use help of experienced mushroom hunters to better predict eadibility. We will spend some time seeking bigger suitable datasets in the Internet. If we success finding one, we will use it. Software: python with some libraries, possibly jupyter notebook.

Project should hopefully be done until deadline.

There is a risk that a dataset this small will not be sufficient to meet the goals. As a "workaround", we will publish as it is but with a feature to add more data (discussed earlier). Also, there is a risk that our clusterization goal will fail (since we do clusterization and not classification, which is much simpler). The "workaround" will be as with the first risk.

Things that grow in forest and have mycelia are mushrooms, the project is centered about such mushrooms.

Mushroom hunter/shroomer/gatherer/hunter - a person who is involved with activity described in the project overview. This activity is called mushroom hunting/shrooming/gathering/hunting. Eadible mushroom is safe to eat if processed properly, no important health concerns present. Ineadible means not eadible.

Cap and gill - two parts of mushroom. Properties of cap are: shape, surface, color. Properties of gill are: attachment, spacing, size, color. The mushroom can or not have bruices and may have a type of odor.

Costs and benefits - our project is not for money, but if we need profit, we may consider making our system paid.

Our model should achieve the following: system correctly identifies at-least 80% of eadible mushrooms (true positive rate), while correctly identifying at-least 90% of not eadible mushrooms (false positive rate not more than 10%). The system can at-least vagually classify the mushrooms so that a manual labeler could label all the clusters; so that our system could actually know species of mushrooms (and display them to the users). The system knows the most significant properties to look for to tell eadible and uneadible mushrooms apart (in future, could be used to create entry-level shroomer guide).

Hard to tell how to estimate other variables of success of our models.