Data Science Test

Welcome to the Seedtag Data Science Test. Your goal is to show us how solid you are at unravelling data related problems and designing solutions. This challenge has two parts, both are equally relevant.

Please include all details you might consider necessary, in terms of code and documentation.

Part 0: Theorical Questions

Answer these questions. We expect to receive a pdf-like output.

1. We are benchmarking two landing pages in our web client. Version A got 1000 new visitors of which 34 signed up and version B got 2000 new visitors of which 84 signed up. Which version performs better?
2. Given the dataset (feature1, feature2, label) (-1, -1, 0), (-1, 1, 1), (1, -1, 1), (1, 1, 0), would you prefer to train a Logistic Regression or a Decision Tree Classifier for label prediction? Why?
3. The probability of a pet being a dog is 30%. The probability that a pet dog weighs more than 3 kg is 60%. If a pet is not a dog, the probability of it weighing less than or equal to 3 kg is 70%. What is the probability of a pet being a dog if it weighs more than 3 kg?
4. We are solving a classification problem using a Decision Tree. When increasing the depth of the tree the accuracy improves until it stagnates at 90%. Why is this an unexpected behavior? Can you offer a plausible explanation?
5. What would you change in this python line

- `operationResult: str = 193;`

Part 1: Experiment Design & Data Exploring

Retrieve the dataset here (also in data folder) and describe it to understand it first, then propose your approach and finally implement the analysis. Expose every issue you may find, it is not a code test (although we appreciate pretty coding) but a data exploring test, so do not forget to include results, observations and written decissions. Your conclusions need to be statiscally grounded.

The output should show your line of thoughts and final ideas, so tell us a story! We expect to receive a notebook-like output with a huge amount of markdown, or any other similar format you would prefer.

In Central Park, NYC there is plenty of squirrels and they are tracked! We would like to know more about them and their behaviour so:

- Make a sociological study about squirrels in Central Park in terms of how they interact between them and towards people.
- Squirrels no jokin'. We suspect there are organized squirrel gangs (maybe dealing with nuts), can you confirm/deny this?
- Friends or foes. Is their friendliness random or related to specific factors?