**Tree leave classification**

**Introduction**

We, Isobel smith, Elise Mol and Wendy Nieuwkamer, have created this repository for the final project of the Machine Learning course at Amsterdam University College. The goal of the project is to demonstrate our ability to understand and implement machine learning techniques on a real dataset.

**Research question**

The problem we are aiming to solve is one of the (challenges on Kaggle)[<https://www.kaggle.com/c/leaf-classification>]. The question is whether we can identify trees by looking at simplified images of their leaves. This has many applications in for example, forest documentation and preservation, crop management and herbal medicine. Also, a fast algorithm might make the development of an instant identification app possible which could be used by campers and backpackers.

We will aim to optimise the classifications as used in the [10 classifier showdown](https://www.kaggle.com/jeffd23/leaf-classification/10-classifier-showdown-in-scikit-learn) and tune the parameters.

The specific classifiers that we will look at optimising are Decision Tree classifier, Gaussian, K-neighbours, and either ada-boost or quadratic discriminant analysis.

**Approach**

We will implement two classification algorithms, namely, simple logistic regression and another algorithm. As the dataset is quite small we will be using k-fold cross validation as a testing mechanism.