

# Project Report

Elise Mol, Wendy Nieuwkamer, and Isobel Smith

21 December 2016

## 1 Problem

After three months of learning about several machine learning techniques and classifying algorithms it was time to put our newly gained skills into practice. We decided to try our hand at one of the competitions hosted by Kaggle. The challenge we chose is Leaf Classification; the objective of the competition is to use binary leaf images and their extracted features to predict what species each leaf belongs to. The dataset which was provided includes around 1584 images of leaf specimens; these samples include 99 species. From every image three sets of features are extracted: a shape contiguous descriptor, an interior texture histogram and a fine-scale margin histogram. Every feature is represented as a 64-attribute vector.

## 2 Approach

As the competition started at 30 August 2016, there were already quite some submissions made and kernels active. One of them stood out, namely "10 Classifier Showdown in Scikit-Learn". In this kernel Jeff Delaney compared ten classifiers which are part of the python library Scikit-Learn. He used the default implementations of the following classifiers:

1. K-neighbors
2. Support Vector Machine
3. Nu-Support Vector Machine
4. Decision Tree
5. Random Forest
6. Ada Boost
7. Gradient Boosting
8. Gaussian Naive Bayes

- 9. Linear Discriminant Analysis
- 10. Quadratic Discriminant Analysis

### **3 Choices made and justifications**

Which classifiers you chose ; why? Describe the relevant features of the classifiers.

## 4 Solution

How do we optimize them? k-fold cv, gridsearch

## 5 Evaluation of Approach and Solution

All the curves, the results

## 6 Sources