Plant disease

Report by Aleksandr Makarov & Nika Mikhailava

Task 1. Setting up (0.5 points)

- Link to GitHub: https://github.com/sudo-Cthulhufhtagn/IDS_Plant_disease

Task 2. Business understanding (1 point)

NB! Don't forget to mention your project title and team members at the beginning of the report.

These days agriculture witnessed boom because of the growing population. The population has surpassed 8 billion people and since demand is growing thee supply adjusts accordingly. In many spheres of agriculture cutting edge technologies like drones and AI used. However, they still remain costly and casual small farms don't have access to them. Our goal is to provide those with cheap deployed machine learning model that can diagnose plants' diseases better than average human does. And by average human we will use team members and a couple of friend as a sample group. Success would be considered if it better than 70%.

In order to succeed in the project we will be using the dataset provided by kaggle from here. Also, we will try tensorflow keras library and fastai for classification purposes. Additionally(depending on the amount of free time) we can try making object detection and it might possibly work because of the data format. For hardware we will be using HPC center in delta, most likely uding jupyter notebook, but scheduling training job is also possible.

The deadline we set for ourselves is December 12 (to have a poster with completeresults). Dataset does not have license or authors in the kaggle, but it is based on the plant-village dataset.

List of possible contingencies: jupyter notebook runtime error, jupyter notebook out of memory, jupyter notebook stopped due to maximum runtime limit. And of course, last but not least: procrastination.

Terminology:
- Image classification - Image classification is the task of assigning a label or class to an entire image. Images are expected to have only one class for each image. Image classification models take an image as input and return a prediction about which class the image belongs to.
- CNN - In deep learning, a convolutional neural network (CNN, or ConvNet) is a class of artificial neural network (ANN), most commonly applied to analyze visual imagery CNNs are also known as Shift Invariant or Space Invariant Artificial Neural Networks (SIANN), based on the shared-weight architecture of the convolution kernels or filters that slide along input features and provide translation-equivariant responses known as feature maps.
- Kernel - a matrix, which is slid across the image and multiplied with the input such that the output is enhanced in a certain desirable manner.
- Object detection - a computer technology related to computer vision and image processing that deals with detecting instances of semantic objects of a certain class (such as humans, buildings, or cars) in digital images and videos.

Cost of the project: free, using university infrastructure or google collab. However, when it would hypothetically come to deployment, the cost of using something like aws lambda will be high because the resulting model won't be small. The prices for different services vary according to load, but we have not made customer survey yet.

Data-mining goals: searching kaggle for something plants-disease related. Downloading it and reformatting for the format needed.

Task 3. Data understanding (2 points)

The report of task 3 should be 400-800 words.

Data was gathered by third partry and published in kaggle with augmentations. Data is confirmed to be available(downloaded and checked). Data contains images of healthy and diseased plants, each located in class folder. Data seems to be high quality, tight images of leaves, though sometimes augmentations look too drastic. The classes of healthy and deseased plants:

Task 4. Planning your project (0.5 points)

1) Data exploration(3 hours). Visualize images and show part of them to the test group(to train them a bit).
2) Try tensorflow keras library for classification purposes. (20 hours including introductory course to tensorflow)
3) Try fastai for classification purposes. (20 hours including introductory course on fastai)
4) Evaluate performance and compare it to humans(non random group of 5 people). (3 hours)
5) Poster creation (4 hours): creating the figures, comparing the models and drawing meaningful conclusion.
6(bonus))   We can try making object detection, maybe, if the time amount of time is not sufficient.(20 hours)

Amount of hours is specified for a team work. We don't distribute the tasks.