Submission from Ilir Hajrullahu (7813401), Lily Djami (7478862), Numan Tok (6927093)

Requirements document for ML-Team

Questions for software engineering teams

General questions:

- What dataset are you going to use?
- What is the expected input?
- How accurate should the model be? → Are there outputs that would be discarded at a specific limit (no result for the user)?
- What is the output that you want, what do you want to predict?
- What is the desired speed of the output creation?
- How do you judge if a predicted output is good or bad?

For nutrition group:

- Is there enough image data, or would simulated data be needed?
- Would the model be used to only detect food from a specific cuisine (Asian/Western/etc.)?
- Will there be multiple objects (different fruits) in the picture or just one? Do you want to detect them one after another or as a group?

For text group:

- Should the language be restricted to English?
- How do you judge the output? How do you know if the summary has the most important words of the long text?
- How long should the summary be (2-3 sentences?)?

For Spotify group:

- How do you judge the created playlist if it is accurate or not?
- Do you want a feedback loop to the model?
- Which features do you want to use for the prediction? Length of song, mood, …?
- Do we have access to the Blend feature functions to possibly build on? (e.g. in the Spotify API)
- Do we also have access to functions of the "Recommended - Based on the content in this playlist" feature in order to use it in our code?
- Is there a group size limit or other restrictions?

We were not sure if we should also provide a VDI Machine Vision document and didn't get feedback from the other groups in order to complete the VDI document. But we also worked out some general requirements for the Nutrition-detection project below.

Each point is divided into 3 for each of the projects (nutrition detection [a], Spotify [b], and text summary [c]). If a point holds true for all projects, it is not further subdivided.

1. Task:

1.1 Type of task
   a. Detection, shape recognition and classification of fruit and vegetables.
   b. Recommendation of a playlist consisting of songs
   c. Summary of an article, as well as fake-news detection

1.2 Objective of testing
   a. Customer benefit, health benefits (time and resource saving)
   b. Customer benefit, time benefit
   c. Customer benefit, time saving

1.3 Detailed description of the task
   a. Detect objects (fruits and vegetables) placed on a surface.(other ingredients?)
   b. Recommend songs from a group of persons based on historic data of the persons
   c. Given a text (i.e. a news article), a summary is created as well as a rating of how reliable the news is.

1.4 Previous procedure
   a. Placement of the fruits/vegetables on a surface? What surface (white,dark,…)?
   b. Gathering of the past data of the persons?

1.5 Characteristics to be validated
   a. Shape, color and features of the fruits and vegetables
   b. Mood of past songs, length?
   c. Key words in the summary, length of summary.

1.6 Objects for validation
   a. Reference data with pictures of fruit and vegetables
   b. Past songs?
   c. Generated summaries

1.7 Validation procedure
   a. Will the model be validated?
   b. Will the model be validated and how?
   c. What is a good summary?

1.8 Evaluation of measurement and tests result
   a. Evaluation with a set of specified pictures which are golden samples?
   b. How do we evaluate if the created recommendation fits the group?
   c. What are the guidelines of a good summary?

1.9 Temporal requirements

What is the required speed of prediction and response time from the model?

1.10 Required availability

When should the ML Model be available to you? All the time? What is a tolerable downfall time for you?

1.11 Special requirements
   a. Accuracy of 80%? What accuracy is satisfactory for you?
   b. What is an acceptable accuracy for you?

2 Test Object

2.1	Range of Types	a. Liquid food? Separated components or mixed components? Food from specific countries? Food from restaurants? b. c. What different types of articles should be fed into the ML system? Only news? Different kinds of news?
2.2	Macroscopic shape and dimensions	a. Will there be a size reference? Will there be standard plates/bowls/tablets etc.?
2.3	Microscopic shape and optical properties	
2.4	Temperature	Shouldn't be relevant
2.5	Variance of object properties	a.How to deal with nutrition variances? (Depending on the data source or sort of component, a component may have different nutrition values)
2.5.1	Preliminary process	c. Is/should the text data be preprocessed?
2.5.2	Object contamination	a.How to deal with added components that can't be seen in a picture? Will there be non-food objects that are on top of the food? (Spoon, knife, …)
2.5.3	Mechanical object stability	
2.5.4	Thermal object stability	Are there differences in cold or warm food?
2.5.5	Temporal object stability	Does the nutrition detection depend on the freshness of the food?

3. Scene (only for nutrition detection)

3.1 Positioning

a. The scene is food ingredients displayed on plates/containers on a table

3.2. Number of objects
   a. There can be multiple ingredients at a time (?)
3.3 Background
   a. The background should be as neutral as possible and with as little distraction as possible.
3.4. Movement
   a. There should not be movement in the picture. Picture should be clear and not blurry.
3.5. Lightning
   a. The picture should be well-lit, i.e. using natural light or taken in a well-lit room.

4. Process Integration

4.1 Use of data (from image processing system)
   a. Images taken by the camera will be used to determine the ingredients.
   b. Use big datasets to train model for recommendation
   c. Datasets will be used to train text summary generation, as well as fake-news detection
4.2. Inputs and outputs (of image processing system)
   a. The input of the ML backend is the image(s?). The output of the ML backend is a selection of recipes that uses the ingredients in the picture (?)
   b. Input is large dataset that the software team gives us? What are the input features? What is the output that the software team wants?
   c. The input of the ML backend is news text. The output is a short, 1-2 line summary, as well as fake-news detection
4.3. Mechanical interfaces
   a. Phone camera will mainly be used to take pictures
   b. None
   c. None
4.4. Operating modes
The ML backend will be operated via API calls
(Which specific functions would be needed?)

5. Human-machine interface

Other than providing API calls, the machine learning team will not be providing user interfaces for the ML backends.

6. Miscellaneous

6.1 Acceptance
Integration with backend and frontend and working end-to-end test with the specified accuracy reached?
6.2 Installation and commissioning
Test runs, human inspection for false classifications?
6.3 Training
Introduction to the system for the software engineering group?

6.4 Documentation
Electronically
6.5 Maintenance
Define maintenance responsibilities, who fixes things when they break?
6.6 Stability and monitoring
Setup monitoring of each part of the system and check availability of the system (API calls work, API requests get a response, etc.)?