Technische Universitaet Berlin

Fakultaet IV - Elektrotechnik und Informatik
Fachgebiet AOT
Prof. Dr. Sahin Albayrak

Bachelorthesis proposal

An intuitive user interface for the automated machine learning project CODA

Robin Ruth

Matrikel–Nummer 316672

Betreuer   Researcher Christian Geissler   Dipl.-Inform.

Inhaltsverzeichnis

1 Motivation

With the ongoing digitalization and the generation of big data in most aspects of society, data driven approaches to every day problems become more and more viable. One way to evaluate these large datasets is with a machine learning approach. Machine learning, though powerful, has its own challenges. Not every machine learning approach is applicable for every problem, not every algorithm will perform in reasonable time on every data, the list goes on.

A lot of intimate knowledge and intuition about machine learning algorithms is needed to select a good approach to a given problem. Knowledge and intuition that is only developed in few experts. Even then, to optimize for a given problem still requires a lot of time, as most of the time the best hyperparameter settings can only be found by trial and error.

To automate this process, GT-ARC (German Turkish Advanced Research Center for ICT) has started CODA, a fundamental research project in algorithm selection and hyperparameter optimization. [Siv]

This project creates an interface to the CODA project, allowing machine learning enthusiasts and specialists to use the developed solutions and giving the team an interface to demonstrate CODAs capabilities.

The provided user interface may be used in demonstrations as well as by persons with a background in machine learning. The focus is set on a clean workflow, good user guidance and a visual appealing display.

2 Objective

The overall objective is to create a web application and a supporting backend to interface with the CODA backend.

Architecture, languages and technologies will not be predefined. Selecting these is part of this project.

2.1 Architecture

The architecture has to support several use cases:

- Demonstration startup

  It should be easy to spin up a clean instance with empty databases for demonstration purposes. Sensible defaults should apply, to make such an instance immediately usable even by an untrained individual.

- Configurability

  Changing the CODA Backend should be possible without recompile.

- Docker deployment

  The System has to run in docker containers on linux.

Some common characteristics of web applications are not relevant in this project:

- User accounts

  All Users will operate on the same data. There will not be any user accounts. Any data uploaded is visible to all other users.

- Login

  The UI will be open and not contain any login features or any other form of access control.

- Responsiveness

  The UI will not be responsiv, mobile usage will not be the focus. A resolution of at least 1024x768 will be assumed.

Coda will run on a separate machine, connected via Kafka-REST-Api. Kafka access is not part of this project. An encapsulation library written in Java will be provided.

2.2 Workflows

Several workflows will be supported:

- Order creation

  An order can be created. This includes the selection of all values needed to create an order, classifier, metric, test dataset and validation method with their respective parameters.

  After an order is created, it will be split in tasks and put into a queue for processing. This includes persisting them to have a keep them in data on restart.

- Test data

  While creating an order, the user may select from a pool of test data but may also chose to upload their own sets. Uploaded files have to be redirected to CODA to allow usage in the evaluation process.

  The data will not be validated on upload. Validation occurs only within CODA.

- Order overview

  View all orders, with their tasks and allows them to be paused or continued if not finished. An instance will connect to a single backend. Changing the connected backend should only occur if the order list is empty.

- Order evaluation

  Display an evaluation page for order and task results.

It is not possible to access multiple CODA backends at the same time.

All evaluation logic or machine learning support will be provided by CODA. As such this project is dependent on the data returned from CODA. Neither does it contain an error logic. If an error occurs in CODA, no help will be provided besides showing the error message that comes from CODA. Correct configuration and service setup will be assumed.

2.3 Evaluation

To evaluate the usability and user interface decisions, a think-aloud interview will be conducted. This method helps to understand the expectations a user has at a given point in a workflow and shows if the user is able to use the application in a successful manner without prior knowledge of the program or if help is needed.

The results of the think aloud interview will be discussed in this thesis, but not influence the actual development. Ideally a think aloud interview would be followed by a secondary development block to address the found factors. Instead the results will be collected for possible future projects.

The think aloud interview is not supposed to be a statistical evaluation, the number of participants does not have to be statistically significant.

3 Work packages

3.1 Introduction

Define the overall goal of this project.

3.2 Basis

Introduce relevant concepts and definitions of Machine-Learning and usability design in general and CODA in particular.

3.3 Requirement definition

- Specify requirements and usecases, feedback

  The general requirements have already been discussed. Specifying them and understanding the usecases for the final product will help understand the needed solution.

  A feedback loop will help find and solve misunderstandings.

3.4 Solution approcach

- Workflowdesign

  Design workflows from an usability point of view.

- UI Design

  Create a user interface design. Start with a design sketch for feedback and create a final design from that.

- Architektur

  Find a fitting architecture for the defined problem.

- Identify standard problems

  Many problems in IT are the same in most programms. Recognizing these and finding fitting solutions will speed up the development process and raise the overall code quality.

3.5 State of the art

- Analysing projects with a similar focus on ai and data analysis.

  Analysing other programs with the same overall purposes can help find good solutions and may also help identify pitfalls.

- Find common problems and standard solutions.

  Selecting the right languages and libraries will help circumvent problems and will reduce the workload to the parts that are actually unique in this project.

3.6 Implementation

The general setup is very similar in all web programs and software projects. CODA connection, Docker setup and the deployment are requirements that are already defined.

- Basic setup

- Order creation

- Data persistence

- Basic user interface

- Test data upload

- Order overview page

- Integrating Kafka access library / CODA connection

- Configuration page

- Order evaluation

- UI optimization

- Bug fixing

- Docker setup

- Final deployment on DAI server

3.7 Evaluation

- Create a questionnaire.

  Create questions and tasks for the interviewed. They should show if a new user is capable of navigating the application, to identify pain points and if they can complete a given objective.

- Find a testgroup for think aloud interview

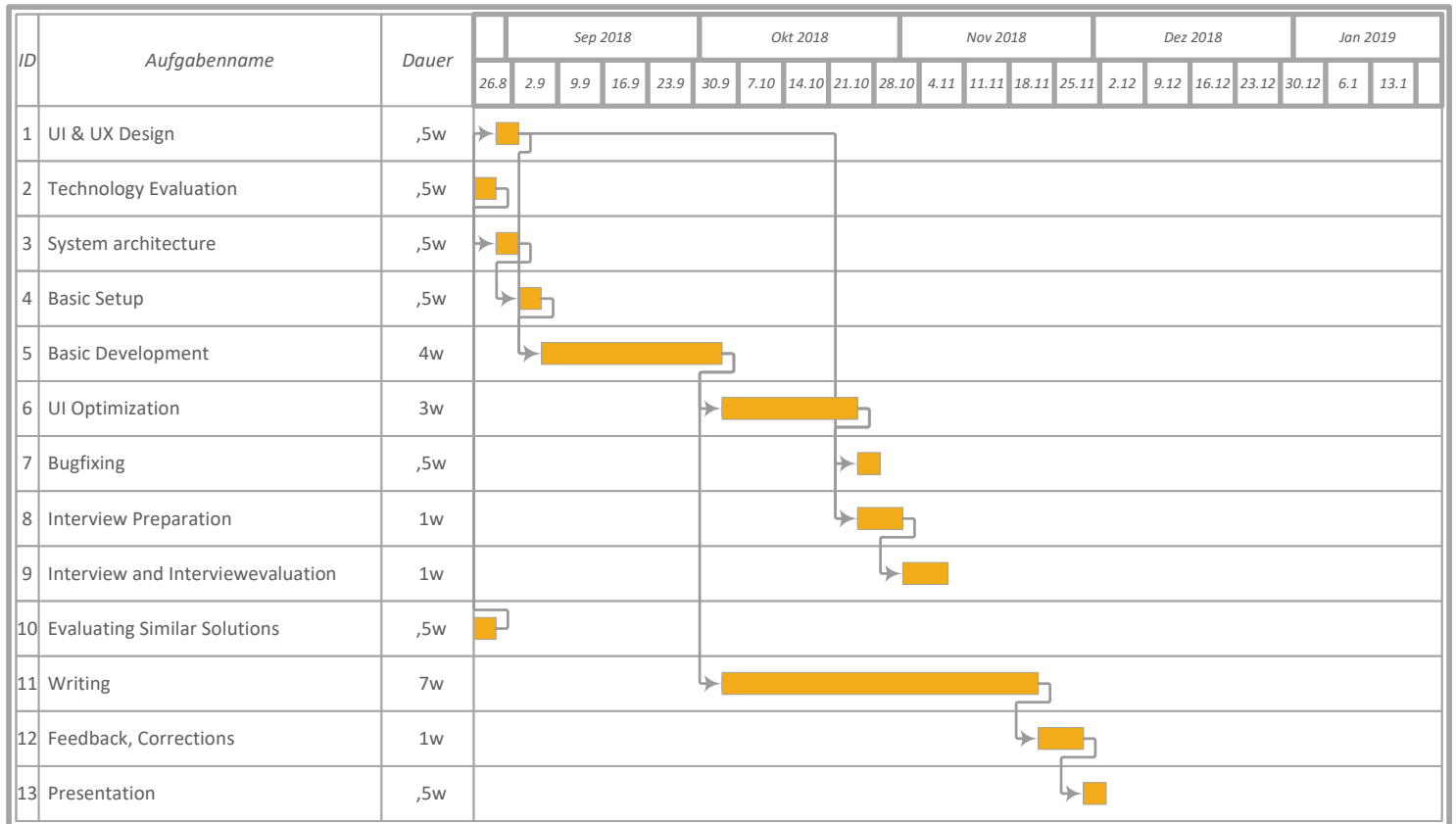 Select a set of 2-4 people of possible futute users to conduct the interview with.

- Evaluate interview

  Where are problem points in the program, where there objectives the interviewed could not solve.

3.8 Conclusion and Finalization

Finalize the project, discuss possible future changes on the basis of the think aloud interview.

4 Schedule

ID	Aufgabenname	Dauer
1	UI & UX Design	,5w
2	Technology Evaluation	,5w
3	System architecture	,5w
4	Basic Setup	,5w
5	Basic Development	4w
6	UI Optimization	3w
7	Bugfixing	,5w
8	Interview Preparation	1w
9	Interview and Interviewevaluation	1w
10	Evaluating Similar Solutions	,5w
11	Writing	7w
12	Feedback, Corrections	1w
13	Presentation	,5w

5 Organizational

- Language of this Bachelor thesis is english.

- The thesis will be written with pdflatex.

- Choosing programming languages and technologies are not defined and part of the development process.

- Supervisors is Christian Geißler

- Evaluators are Prof. Dr. Albayrak and Prof. Kao

6 Appendix

Literatur

[Siv]   Dr. Fikret Sivrikaya. *Automatisierung von Machine Learning Workflows, CO-DA Projekt-Steckbrief*. URL: http : / / www . gt – arc . com / wp – content / uploads/2017/08/Projekt_GT-ARC_CODA_de.pdf.