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

ML Team - Assignment for 19.11.2022

Summary about the video “Fostering interdisciplinary teams”

Prof. Kaestner made a point regarding the structure of the project team and said that the structure should be aligned to the code structure. A possible conflict point between the Machine Learning and Software Engineering parts of the project may be data accuracy and system efficiency. We must be aware that we may not reach optimal results in both of these categories and we may have to settle for a threshold on both topics when discussing this in the project. Another problem may be not knowing what the next steps are and focusing on the least important parts of the project. Prof. Kaestner mentioned several problems that could occur in group work. These problems are group thinking, social loafing, and lack of motivation. Group thinking is when everyone in the group agrees on something without any further discussion. This is bad for the work quality of the team, as no one questions the decisions that are made, or checks the quality of the work within the team. Conflicts should be tolerated to a certain level, because they may create new insight on the project that the group didn’t think about earlier. Social loafing is when the team has low work motivation because everyone has the mentality that they are in a team and someone else could pick up the work.

Knowing these issues, we come up with the following strategies for our team:

- Avoid Group thinking
  → We could avoid this by letting everyone think about the discussion topic beforehand (e.g. planning poker)
  → Look at code from other people without trusting that it is good
- Fostering interdisciplinary teams
  → We as ML Team could try to split into different teams
- Regular meetings to check on progress
  → This helps with team motivation and accountability, avoiding social loafing.
  → Everybody knows what everybody is doing and where the problems are
  → Encourages a culture of open conflicts
- Minimize Process Costs
  → Ask other teams to declare a team speaker
  → Shorten the communication between teams by dividing the project into modules
- Task prioritization
  → Sprint planning for the next 2-3 weeks
  → Focus on the most important things
  → Deliver fast results

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

Summary about the paper “Data Scientists in Software Teams”

Modern software teams utilize the power of data to make decisions for their products. The trend for the topics of data scientists in software projects is moving towards analyzing user/customer behavior. Data Science is also used for analyzing software quality in large companies. There are a lot of people who started on other branches (Software Engineering) and transitioned to the Data Science field. These people are called “Moonlighters” by the authors. They have skills in software engineering and data scientists and are sought after.

The paper mentions several challenges that data scientists face. Good quality data might be difficult to obtain. One might need to combine data from different sources, which can be difficult. Data quality might be insufficient due to poor sampling or collection methodology. Furthermore, feature engineering is non-trivial. Lack of understanding of the data due to bad documentation or naming inconsistency can influence the analysis. Scalability of machine learning models may be expensive when processing large datasets.

Best practices:

- Good understanding of statistics and machine learning
- Set clear goal for the project
- Using homogenous tools and standardized data
- Making sure the data makes sense in the context of the problem
- Repeat the same procedure multiple times to ensure correctness of results
- Create own data / simulations
- Understand the data. When the math is right, the answer is not necessarily right or it may not make sense.
- Recognize biased data and always question the data and your assumptions.

Summary about the kickoff ML Team Meeting from 12.11.2022

Our Goal:

- Challenge the other teams to be able to say what is their
    - Dataset
    - Assumptions
    - Context in which the system works
- Form certain practicalities which are oriented to the best practices summarized above
- Maybe a multimodal way approach may be appropriate for the projects
- Consulting team

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

Nutrition Detection:

- It is the least well defined topic
- One of the problems is “amount of food in dish” and the classification
- What is the data set, what are the assumptions (these questions are for the software engineering team)
- Shaping the constraints is necessary
- Each food may be a composition/mixture
- Data needed: images of food and nutritional database
- Lily may help as machine learning consultant
- Numan may help as a user bridge between this group and the ML group
- Does prior work exist or references? CurryAI (https://analyticsindiamag.com/how-i-created-curryai-a-computer-vision-aided-indian-food-nutrition-calculator/)

Spotify Group Playlist:

- A possible way to solve the matching song problem could be to take pairwise recommendation and adjust it for more people
- It may be a “psychological problem” regarding the group preference and recommendation
- Reinforcement learning can be considered (Feedback loop with votes from users) but time may not be sufficient
- Numan may help as a technical consultant
- Ilir may help as a user bridge between this group and the ML group
- There is an existing project proposal on this topic which may help to form ideas: AMGPRST (AMGPRST · GitHub)

News Summary:

- Create a good summary - what is a good summary?
- Extract important information from the text
- Challenge may be using GPT-3 on a mobile phone
- Put information into well written text
- Ilir may help as a technical consultant
- Lily may help as a user bridge between this group and the ML group
- Existing project: News Summarization and Evaluation in the Era of GPT-3 (https://arxiv.org/abs/2209.12356)