Stay at home — Are Germany's greenhouse gas emissions increasing or decreasing

Proposal of Group 5

Robert Annuth	Laura Pilger	Mohamed Badawy	Natalia Voronova	Utku Ayvaz
03723414	03735972	03722512	03691009	03690266

Noah Binh Nguyen	Mariem Nouicer	Samra Hodzic	Muhammad Zubair Jamil
03683505	03687720	03723863	03693920

26/05/2020

Motivation

Besides the current Covid-19 issue the population of the whole world is facing the climate change. Almost all human activities cause or lead to greenhouse gas emissions, which drive up the temperature. Extreme weather and melting polar ice are other possible effects. Consequently it is highly relevant to identify economic sectors and human behavior with the potential to decrease emissions.

Current research regarding this topic is mostly based on mathematical models and can not be verified by actual data. Due to the restrictions of the Covid-19 crisis the human behavior has transformed, which creates ground truth data and the possibility to conclude if we can decease greenhouse gas emissions by maintaining behavior, caused by restrictions during the crisis.

---

1 Project Description

The current research evaluating human behavior, decreasing green house gas emissions, is always based on the results from areas where the behavior is present. The results then are transfered to different areas, leading to predictions, not considering the local circumstances, which could be completely different. The Covid-19 issue is creating actual data to analyse the impact of human behavior on greenhouse gas emissions.

On the one hand economic sectors can be analysed and potential factors derived, which lead to decreasing emissions. On the other hand the general question “What can we do to decrease greenhouse gas emissions?”, which refers to the actions every human can take, can be answered based on actual data.

Research Question

Is it possible to decrease Germany’s greenhouse gas emissions by maintaining behavior of the Covid-19 issues?
This research analyses which human behavior, during the Covid-19 crisis, is increasing or decreasing greenhouse gas emissions in the scope of Germany. The increase or decrease is measured in respect to the time before the crisis.

Goals

Our goal is to find human behavior, decreasing the total greenhouse gas emissions. We think that some of our Covid-19 behavior is strongly influencing the total emissions. Hence, it is interesting to find out how the human behavior correlates with the decrease in emissions of each economic sector. Besides we are curious to find out if there is human behavior which is significantly decreasing the total emissions and could be maintained beyond the crisis to archive the requirements of the climate agreement.

Approaches

The following list is containing topics providing information to answer the research question.

- The German “Umweltbundesamt” is providing actual CO2 data
- Real time monitoring of greenhouse gases with satellite pictures

- Information from economic sectors can be used as additional predictor — for example the electricity sector
- The consumption of resources (oil, coal,...)
- Stock market information
- News about restrictions

The German office for environment is sharing the data they are using for their reports. Currently no report regarding greenhouse gas emissions for 2020 is available, nevertheless they provide the data files. This way we can acquire actual ground truth data to evaluate our results. Another interesting approach is to evaluate satellite pictures, as it is possible to extract information about the amount of greenhouse gas in the air. The correlation between the human behavior and economic sectors depends on detailed information how each sector is performing during the crisis, these are extracted form the remaining data sources.

2 Work packages

Our group found a detailed explanation how to structure larger machine learning project and we agreed to follow this guide. (link)
Nevertheless, the guide is mostly concluding what we discussed inside the breakout rooms during the second lecture.

- **Planning and project setup**
  - Define the task and scope out requirements
  - Determine project feasibility
  - Discuss general model tradeoffs (accuracy vs speed)
  - Set up project codebase
- **Data collection and labeling**
  - Define ground truth (create labeling documentation)
  - Build data ingestion pipeline
  - Validate quality of data
  - Revisit Step 1 and ensure data is sufficient for the task
- **Model exploration**
  - Establish baselines for model performance
  - Start with a simple model using initial data pipeline
  - Over-fit simple model to training data
  - Stay nimble and try many parallel (isolated) ideas during early stages
  - Find SoTA model for your problem domain (if available) and reproduce results, then apply to your dataset as a second baseline
  - Revisit Step 1 and ensure feasibility
  - Revisit Step 2 and ensure data quality is sufficient
- **Model refinement**
  - Perform model-specific optimizations (ie. hyperparameter tuning)
  - Iteratively debug model as complexity is added
  - Perform error analysis to uncover common failure modes
  - Revisit Step 2 for targeted data collection of observed failures
- **Testing and evaluation**
  - Evaluate model on test distribution; understand differences between train and test set distributions (how is “data in the wild” different than what you trained on)

- Revisit model evaluation metric; ensure that this metric drives desirable downstream user behavior
- Write tests for:
  * Input data pipeline
  * Model inference functionality
  * Model inference performance on validation data
  * Explicit scenarios expected in production (model is evaluated on a curated set of observations)

3 Workload distribution

First, we decided to assign specialists to certain tasks (table 1). The specialists have experience to accomplish their task and therefore they are responsible to keep track of the process and improve the work flow. Additionally they are the first person to contact when question regarding their task occur.

Task	Responsible Person	Description
Data Engineer	Samra	Collecting Data, Quality, Preprocessing
Machine Learning Engineer	Mohamed	Model creation
Software Engineer	Mariem	Front End
Project Manager	Zubair	Planning, Documentation
Video	Natalia	

Table 1: Project tasks assigned to specialists

Second, we assigned the work packages discussed in section 2 to group members and specialists. All tasks are containing the subtasks in respect to the work package.

Task	Responsible Person	Specialist
Submit Milestone I	Robert	Project Manager
Submit Milestone II		Project Manager
Submit Milestone III		Project Manager
Submit Milestone IV		Project Manager
Submit Video	Natalia	Project Manager
Update Trello (link)	Laura	Project Manager
Identify/ Collect Data Sources		Data Engineer
Create Front End	Mariem, Robert	Software Engineer
Pre process Data		Machine Learning Engineer
Choose and Create Model	Mohamed, Utku, Noah	Machine Learning Engineer
Pipeline	Zubair, Robert	Project Manager

Table 2: Workload distribution

4 Risk Analysis

Our group tried to identify possible risks during the project phase. The following list contains the results of a short brainstorming.

- Quality of the data
- Not enough data
- No existing data for performance measurement - No ground truth
- We are discussing a complex issue, obtaining clear results can be difficult
- No program for video editing
- Not enough computing power to train the models
- Version control (updates of packages)
- Organisational issues (no personal meetings, big group, conflict potentials)
- Other lectures (overlapping deadlines)
- Research question is too complex

We agreed, that we are facing two risk. One is not to find enough data, which would make it impossible to make good predictions. The other issue is that we could be unable to conclude which human behavior is causing decreased or increased emissions, because the correlation is too complex.
The first issue could be a real problem because besides putting more effort into finding data, there are no solutions. The second issue can be solved with more data or better predictors.

The remaining issues result in additional workload but do not jeopardise the project. However, with the possible data sources mentioned earlier we are confident, that we are well prepared for the project phase.

5 Time Table

We decided to organise our project using a Gantt chart. There we are including all milestones and additional tasks required to fulfill each milestone. Fig. 1 and 2 show a short overview of the current plan.

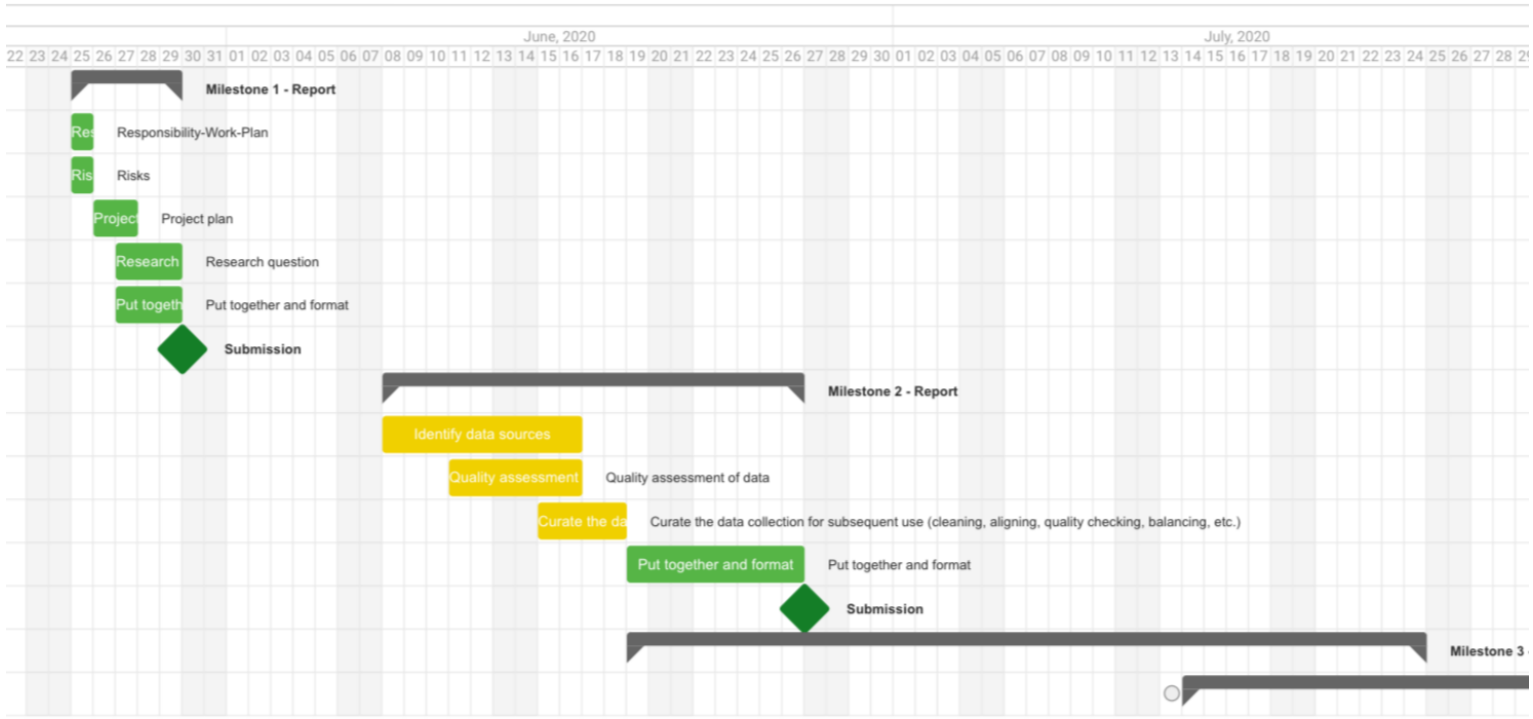

Figure 1: Gantt Project Plan Milestone I and II (26.05.2020)

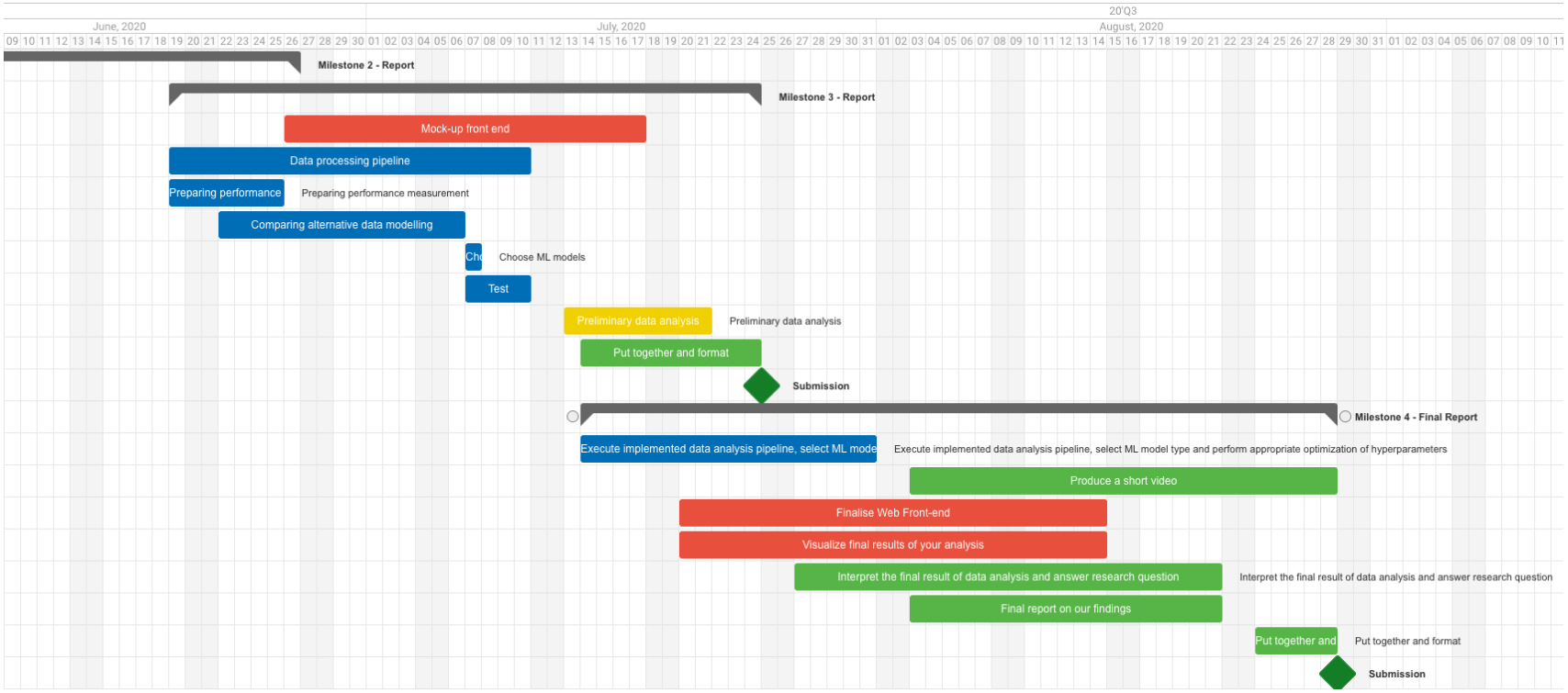

Figure 2: Gantt Project Plan Milestone III and IV (26.05.2020)