Using Machine Learning to Predict the Effects of the COVID-19 Pandemic on Reaching the EU Climate Goals for 2030

Proposal - Group 10

Florian Auinger	Florian Butsch	Florian Hölzl	Christoph Miller
03659767	03663488	03659355	03721920
Johannes Gensheimer	Johannes Kiechle	Johannes Gahr	Constantin Nowak
03721520	03735507	03675464	03676376

June 16, 2020

Motivation

We are currently in the middle of the COVID-19 pandemic, the biggest pandemic in the last 100 years. While all of our lives are rapidly affected by it, other issues are overshadowed. The changing climate is one of the most severe challenges for mankind on a long term basis. In recent years, initiatives to fight climate change and global warming have been raising awareness about the topic. There exist different scenarios by a variety of researchers which all point towards the same fact that greenhouse gas emissions have to be reduced to keep global warming in an acceptable range. Germany's goal is to reduce greenhouse gas emissions by 55 Percent until 2030 compared to 1990. However, there are already forecasts that predict the failure of this objective.

The effects of the COVID-19 pandemic are easy to observe. Many countries have decided on lockdowns for both, the economy and individuals in order to slow down the spread of the virus. Consequences are among others a drastic decrease in industrial production due to closed factories and reduced demand as well as travel bans and even restrictions on all non-obligatory personal movement. Currently, a noticeable decrease in greenhouse gas emissions can already be measured, but so far, there is no certain forecast on how the numbers will develop during the time of this crisis and thereafter. In this project, we are going to try to tackle this problem and provide a range of possible scenarios on how Germany's emissions are going to develop further down the crisis and when the country goes back to a normal daily routine. These resulting scenarios and their significance will then allow us to discuss Germany's adherence to the EU Climate Goals for 2030.

---

1 Project Description

Recent data has shown, that the COVID-19 pandemic and the associated public lockdown resulted not only in a decrease in mobility and consumption but also in a reduction of greenhouse gas emissions in Germany. The scope of this project is to analyze these effects and possible future developments after the crisis in Germany. Machine learning techniques are used to predict the short- and long-term change in emissions based on currently available data. Later, the results are used and put into context of the EU Climate Goals for 2030 in order to better evaluate the impact of the crisis on a bigger scale. Germany, as the biggest economy and with the largest population of any member country in the European Union, is used as a representative to highlight the consequences of the crisis and the associated measures imposed on the economy and the population for our current approach on combating climate change.

Research Question

Which effects will the COVID-19 pandemic have on Germany's greenhouse gas emissions in the context of reaching the EU Climate Goals for 2030?

Goals

The goal of this project is to accurately predict the future greenhouse gas emissions with different models and visualize the different outcomes within an interactive user interface. Within the scope of this work we will investigate if and to which extend this global pandemic will help Germany to reach its climate goals. Furthermore, it will be possible to simulate different future scenarios depending on a range of different indicators affecting emission. We will provide a user friendly visualization interface showcasing the results and informing the user about the findings of this project.

Approaches

In order to measure the impact of the COVID-19 pandemic, we develop a data-driven ground truth for greenhouse gas emissions by using past measurements to model the general trend in Germany. Afterwards, we analyze the currently available data on carbon dioxide, NOx and other climate contributors during the crisis to find the best suited prediction models and compare their results. In addition to direct emission measurements, possible key indicators, such as traffic volume, stock market indices and weather data are evaluated to help improve our prediction models. In the end, the findings are put into context by highlighting the assumptions necessary for each model and by discussing the advantages and disadvantages of each approach.

2 Work packages

The work packages relate to the project plan and the Milestones. Additionally we identified packages that support the given requirements. Each package is divided in subpackages which can be assigned to contributors of this project. This work package plan is the basis for the following workload distribution and time table.

For all milestones documentation of the work and preparing the deliverable product is part of the work packages and not specifically listed again.

- **Package 1: Milestone 1 - Research Question and Project Proposal**
  - **Research question**
  - **Project plan**
  - **Workload splitting**
  - **Potential risks**
- **Package 2: Milestone 2 - Data Collection Pipeline**
  - **Data identification** identify data areas, identify data sources
  - **Database organization**
  - **Data collection** download data, build web crawler
  - **Data preparation** judgment/quality checking, cleaning, aligning, quality checking, balancing
  - **Integration to pipeline** create interface organization
- **Package 3: Milestone 3 - Data Analysis Pipeline**
  - **Data analysis pipeline organization**
  - **Data preprocessing** cleaning data, data transformation, dimensions reduction, Splitting data into different sets
  - **Identify machine learning methods**
  - **Implementation** Model 1, Model 2, Model 3
  - **Evaluation** measuring and assessing the quality or relative performance
  - **Visualization** standardized presentation
  - **Front-end design**
- **Package 4: Milestone 4 - Final Result**
  - **ML method selection and optimization** ML method selection, hyper parameter optimization
  - **Web front-end** visualization of implementation
  - **Report** data basis, data model used, discussion of results, answer to research question
  - **Video production** concept, filming/animating, film editing

3 Workload distribution

First, we define the main persons responsible for different areas, so that they keep a certain overview.

General responsibility	**Person**
Organization	Florian Hölzl - FH
Data	Christoph Miller - CM
Pre-Processing	Johannes Gahr - JGa
Analysis and Prediction	Johannes Gensheimer - JGe
Deliverables	Florian Butsch - FB
GUI	Johannes Kiechle - JK
Documentation	Florian Auinger - FA
Video	Constantin Nowak - CN

However, since certain areas build on others, some people would be left idle. To avoid this, the work is distributed from milesteone to milestone to ensure that each team member has something to work on at all times. Since we are a rather homogeneous team this kind of work distributions allows everyone to work on different tasks and have therefore a higher learning outcome. The subsequent tasks start at the specific milestones, but are sometimes processed iteratively throughout the entire time of the project. To enable the integration of different subtasks, we define an organizational task for each milestone. This task takes care of the infrastructure to bring program parts together.

Milestone 1 - Research Question Project Proposal		
Tasks	**Subtasks**	**Persons**
Research question		FB, FH
Project plan		CM, JGe
Workload splitting		JGa, CN
Potential risks		FA, JK

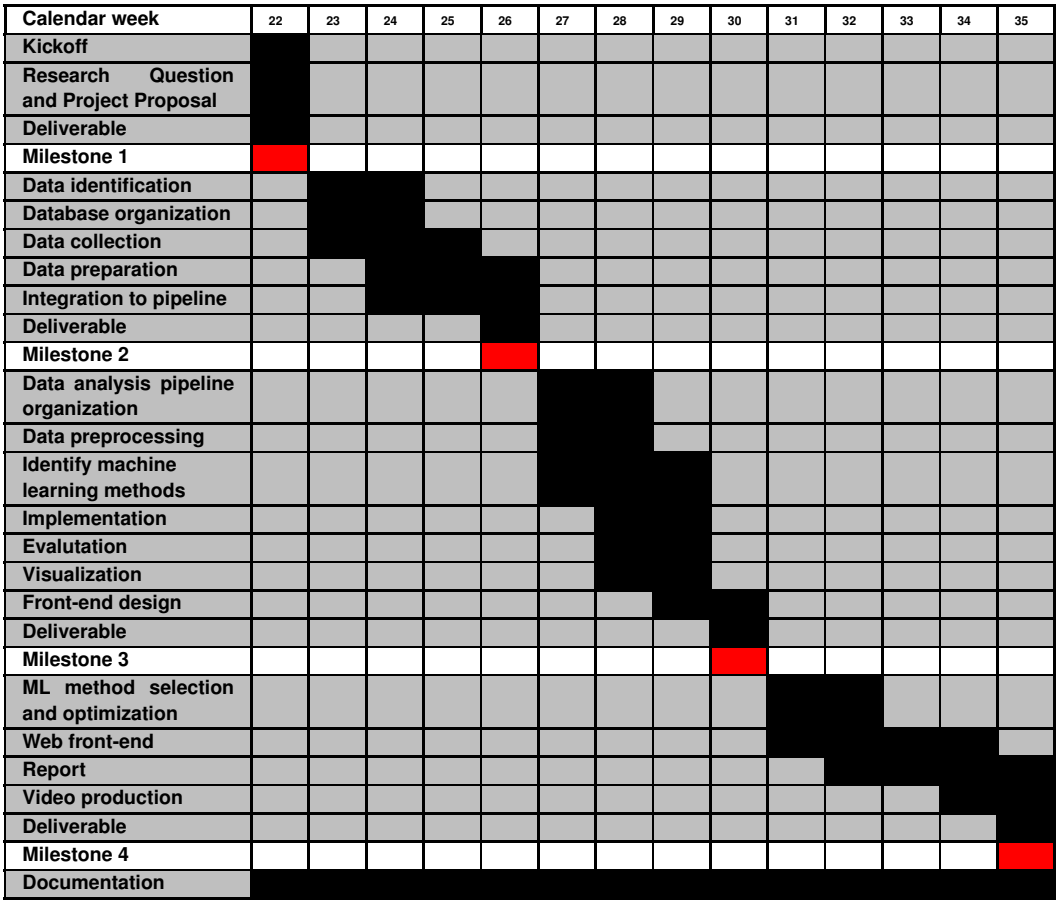
Milestone 2 - Data Collection Pipeline		
Tasks	**Subtasks**	**Persons**
Data identification	Identify data areas	All
	Identify data sources	CM, CN
Database organization		FH, FB, CM, JGa
Data collection	Download data	FA, JGe
	Build web crawler	FB, JGe, JGa
Data preparation	Judgment/ Quality checking	All
	Cleaning, aligning, quality checking, balancing	JGa, JK
Integration to pipeline	Create interface organization	JK, FH
Documentation	Describing data sources and measures for preparation of the data	All
Deliverable	Milestone report	All
	Pointer to data set	FH, FA

Milestone 3 - Data Analysis Pipeline		
Tasks	**Subtasks**	**Persons**
Data analysis pipeline organization		FH, FA
Data preprocessing	Cleaning data	JGa, CM, JGe
	Data transformation	JGa, CM, JGe
	Dimensions reduction	JGa, CM, JGe
	Splitting data into different sets	JGa, CM, JGe
Identify machine learning methods		All
Implementation	Model 1	TBD
	Model 2	TBD
	Model 3	TBD
Evaluation	Measuring and assessing the quality or relative performance	JGe, FB
Visualization	Standardized presentation	CN, FA
Front-end design		CN, FA
Documentation		All
Deliverable	Milestone report	All
	Mock up of front-end	CN, FA
	Software running test set	FH, CM

Milestone 4 - Final Result		
Tasks	**Subtasks**	**Persons**
ML method selection & optimization	ML method selection	FB, JGe
	Hyper parameter optimization	All
Web front-end	Visualization of implementation	FH, CM, CN
Documentation		All
Report	Data basis	CM, TBD
	Data model used	JGe, TBD
	Discussion of results	All
	Answer to research question	FB, TBD
Video production	Concept	CN, JGa
	Filming/ Animating	CN, JGa
	Film editing	CN, FH
Deliverable	Report	All
	Video	CN, JGa, FH
	Software (Python implementation and web front-end)	All

4 Time Table

The time table includes the defined work packages. The chronological order results from the given submission dates for die Milestone documents. Submission dates are marked in red and have to be done till Friday of the corresponding weeks. Every week on Thursday the group meets via Zoom. If necessary additional group or sub-group meetings can be arranged.



5 Risk Analysis

Within a successful project management the risk analysis is an enormously important part and is therefore one of the key components to reduce the occurrence of bad surprises and uncertainties on the way to reaching the desired goal.

A large number of IT projects do not fail due to a lack of technology, but simply because of problems in the teams, their communication or missing/incorrect management. Many concerns therefore relate both to the organization of the project itself and to its planning and implementation.

The following table is an excerpt of the performed risk analysis and does not cover all possible contingencies, but rather provides an analysis combined with a summary of the most common and threatening risks that occur based on our experience and imagination. Although conducting an a priori risk analysis is inevitability, it does not guarantee properly applied risk management during the project process. It rather emphasizes the key factors to which particular attention should be dedicated to.

Risks in descending order of severity	**Countermeasures to be applied**
Appropriate time investment among all team members due to attendance of other classes and obligations	Clarify commitment in advance in Team Canvas, equal sharing of the workload, precise distribution of tasks, assign responsibilities
Organization of group cooperation due to the considerable number of members	Usage of cloud-based collaboration tools (e.g. Miro Whiteboards), on demand communication via web-services (e.g. WhatsApp, Zoom), fixed weekly team-meeting through Zoom
Project time delay due to work package dependencies	Prior clarification of dependencies, synchronization and completion of work packages according to project time table
No ability for the group to meet in person, which handicaps discussions and bonding among team members	Frequent team / sub-team communication according to the work packages using online services (e.g. Zoom)
Availability and access of data	Evaluation through early investigations (i.e. evaluated data availability before framing the research question)
Composition and quality of data	Distributed data collection using various data sources, initial verification of data using common-sense
Lack of knowledge / knowledge transfer while implementing machine learning methods	Self-study, attend AMI class, utilization of AMI-class discussion forum, consultation with other groups
Insufficient documentation of the project, including work packages, code, as well as project structure	Assigning person in charge for documentation, management and the deliverables, continuous updating and improvement of the documentation by the whole team