



Coronas impact on climate change via increase in Internet traffic

Proposal

Martin Schuck 03683571	Michael Brandner 03681315	Alexander Griessel 03687564	Aladin Djuhera 03671760
Niklas Landerer 03671815	Felix Montnacher 03689999	Aron Endres 03689152	Maximilian Putz 03682862

Henrique Soares Frutuoso
03682137

May 29, 2020

Motivation

The Coronavirus pandemic and its consequences are the dominating topics in most of the recent discussions and research. Due to the sheer scale of this crisis, it is not at all clear what its impact on society, economy and our environment will be. There have been a number of reports on the positive effect on greenhouse gas emissions as well as air quality. However, the recommendation for social distancing has also lead to a surge in the use of virtual replacements and a rise in Internet traffic. The significant increase in emissions that can be expected from these activities has found less interest in the public so far. Nevertheless, the global emissions of the Internet are of relevant scales and comparable to the airline industry. Furthermore, with streaming and home office there is a trend towards higher data volumes consumed per person. Under normal circumstances, only the servers located in Germany alone use about the same energy as the whole city of Berlin. Because exploring this increase lies in the intersection of our area of expertise and the pressing matter of estimating the consequences of this pandemic, we deem this a very suitable topic to work on. It is also relevant because this increase in the use of digital technologies has long been predicted to happen for the coming years anyway, although not as fast as it is happening now.

1 Project Description

During this project, we want use machine learning approaches to model the current and future greenhouse gas emissions due to Internet traffic in the context of the 2020 Corona pandemic. It is already known that the Internet continually increases in contributing to man made climate change. It is also known that web traffic grows by about 30% on a yearly bases during normal times. Based on recent data about the growth induced by the pandemic, we want to quantify the irregular increase of this contribution to greenhouse gasses. We also want to make the model and its predictions accessible for users and add interaction possibilities.

Research Question

To concisely formulate our project task, we pose the following research question: What will the global quantitative impact of increased Internet traffic due to Corona be on greenhouse gas emissions?

Goals

In the course of this project, we want to learn a model that predicts these greenhouse gas emissions from increased web traffic as accurately as possible. The predictions should be visually accessible with a web interface. Users should also be able to interact with the model by entering different scenarios such as whether there will be a second wave of infections in autumn 2020. The project is finished once the web environment is working, our model has been integrated and the our progress and achievements are documented by a video we intend to produce.

Approaches

We approach our problem via data driven analysis. In the beginning of the project, we plan on collecting as much data as we can find on the energy consumption of data centres and Internet infrastructure as a whole. We will also try to gather information about the current greenhouse gas emissions for power generation as well as their historical development for individual countries. Combined with the data on web traffic and its increase due to COVID-19, we intend to train a variety of machine learning models to predict the increase in emissions during this pandemic. For future emission predictions, we will either rely on forecasts about the course of the pandemic, or try to generate these ourselves. For the web interface, we plan on using the Chart.js framework to create a visually appealing and interactive environment.

2 Work packages

To organize the project, we define the following work packages that closely follow the general process of creating a machine learning model.

- **Package 1: Data collection**
 - **Data compilation:** Find sources, make accessible
 - **Data inspection:** Assess quality, feature selection
- **Package 2: Data preprocessing**
 - **Visualization:** Gain insight on data, create data set
 - **Processing:** Normalize, PCA, scaling, feature completion
- **Package 3: Model design**
 - **Model selection:** Implement different algorithms, compare approaches
 - **Model architecture:** Hyperparameter search, validation
 - **Testing:** Separate model tests
- **Package 4: User webinterface**
 - **Visualization:** Choose framework, select design, implement
 - **User interaction:** Choose different scenarios
- **Package 5: Video creation**
 - **Planning:** Collect ideas, write storyboard
 - **Production:** Capture, cut, music
- **Package 6: Project management**
 - **Communication:** Team coordination, document milestones
 - **Organization:** Manage schedule, progress checking, weekly meetings

Each group member will be assigned multiple, ideally non time conflicting packages in the next section.

3 Workload distribution

We judge data collection to be one of the biggest working packages. However, without data, most of the other packages won't be able to start their work. For this reason, all team members are assigned to the data collection task. In order to avoid a missing sense of responsibility for this task, the research will be divided into subcategories for smaller groups of 2-3 members. Final data extraction will then be done by the data collection team. The subtopics are assigned as follows:

Subtopic	Member 1	Member 2	Member 3
Internet traffic	Maximilian Putz	Felix Montnacher	
Web energy consumption	Martin Schuck	Aladin Djuhera	Henrique Soares Frutuoso
Course of the corona progression	Michael Brandner	Aron Endres	
Electricity mix history	Niklas Landerer	Alexander Griessel	

The other packages will be divided between the team members as follows.

Name	Assigned work
Martin Schuck	Software architecture, Model design, Project management, Data collection
Michael Brandner	Data preprocessing, Model design, Software architecture
Alexander Griessel	Project management, Data preprocessing, Software architecture
Aladin Djuhera	Data preprocessing, Video, Data collection
Niklas Landerer	Web interface, Video, Software architecture
Felix Montnacher	Web interface, Data collection, Data preprocessing
Aron Endres	Data preprocessing, Model design, Video
Maximilian Putz	Web interface, Model design, Data collection
Henrique Soares Frutuoso	Web interface, Data preprocessing, Model design

Changes to the planned schedule might of course occur and resources redistributed accordingly.

4 Time Table

Data collection of course stands at the very beginning of the agenda. This is also required for the second milestone. Apart from defining interfaces the abstract process, data preprocessing can really only start once the data sets are acquired. The same goes for model design which needs the preprocessed data as input. Web development on the other hand can start independently from the model or any data and will therefore run in parallel with these packages. Due to the sequential timeline of the first three, we judge these as the most time consuming tasks. The video creation really only makes sense at the very end of the project when we know the content we can use. Management of course is needed throughout the whole project.

Calendar week	22	23	24	25	26	27	28	29	30	31	32	33	34	35
Milestone 1														
Data collection														
Preprocessing														
Milestone 2														
Model design														
Webinterface														
Milestone 3														
Video creation														
Management														
Milestone 4														

5 Risk Analysis

We identify the lack of data on the Corona pandemic and web electricity consumption as the biggest risk to our project. In case our data sets show a poor quality or essential data is missing for certain countries, we might need to limit our question to for example only western industrialized countries where we expect reliable data to be available. Furthermore, this pandemic produces highly atypical data with a very small sample size. This might lead to a poor model performance. In this case, we plan to further simplify our models and feature spaces. If everything else fails, we can only explain why our approaches did not produce useful predictions. We judge the other risks, although worth considering, as comparably manageable with sufficient commitment from each group member.

Risk	Countermeasures
Bad or missing data	Adapt question
Bad model behavior	Simplify further, worst case: explain reason for failure
Time constraints	Reassignment of member responsibilities
Hardware limitations during training	Google Colab and sufficient hardware from all members
Difficulties with user interaction	Restrict input possibilities
Underperforming group members	Overperforming Nick