



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 30, 2020

Motivation

Recently, the Covid-19 pandemic and its consequences have dominated public discourse and scientific research. It is yet unclear to what extent the sheer scale of the Covid-19 crisis will impact our society, our economy, and the environment. Covid-19 induced reductions to travel and manufacturing industries has led to a number of reports and studies discussing the positive effects of the pandemic on greenhouse gas emissions. However, negative effects of the pandemic on emissions have found little attention. Due to social distancing policies inducing an increased demand for virtual replacements of social and workplace communication, internet traffic has surged. This surge in internet traffic can be expected to increase internet-related emissions, which represent a relevant portion of total global emissions- nearly as high as emissions associated with the airline industry.

As people spend more time at home, they consume more internet-related entertainment in the form of streaming media, and spend more time and data on home-office related tasks. This has led to a trend in increasing data volumes by person. Under normal, pre-Covid circumstances, servers located in Germany required as much energy as the entire city of Berlin in a comparable time-span. Pressing public and policy-maker interest in estimating the consequences of the pandemic, as well as this group's experience and interest in this area make this a suitable and relevant topic to work on. The results of the progress made on this topic are additionally applicable for the future, as the long predicted rise in digital technology volume will have an eventual effect on greenhouse gas emissions.

1 Project Description

The goal of this project is to model current and future greenhouse gas emissions arising from internet traffic in the context of the Covid-19 pandemic, utilizing a machine learning approach. Prior to the outbreak of Covid-19 it was already predicted that increased internet traffic would result in a continuing rise of greenhouse gas emissions. Pre-Covid-19, web traffic experienced a growth of roughly 30% per annum. Based on recent growth data in internet traffic caused by the pandemic, we want to quantify the irregular rise of greenhouse gas contribution. The model and resulting predictions should be made accessible to a wider public, with the help of an interactive web interface.

Research Question

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

Goals

Over the course of this project, we want to generate a model that predicts greenhouse gas emissions from increased internet traffic as accurately as possible. The predictions should be visually and interactively accessible with a web interface. Users should be able to interact with the model by entering various scenarios, such as differing curves for the progression of the pandemic. The project is finished once the generated model is integrated with a working web interface, and the progress and project achievements are video-documented.

Approaches

The problem is approached via data driven analysis. At the start of the project, we plan on collecting any data we can find on the energy consumption of data centres and internet infrastructure or internet traffic. We will also gather information on the current greenhouse gas emissions for power generation, as well as the historical development of greenhouse gas emissions for individual countries. Combined with the data on web traffic and its increase due to COVID-19, a variety of machine learning models will be trained to predict the increase in emissions during the pandemic. Future emission predictions will rely on forecasts about the course of the pandemic, or will be predictively generated based on past data. 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, work packages are defined. The defined work packages closely follow the general process of machine learning projects.

- **Package 1: Data collection**
 - **Data compilation:** Find, download, and compile source data
 - **Data inspection:** Assess data quality for completeness and reliability
- **Package 2: Data preprocessing**
 - **Visualization:** Gain insight on data, create final data set
 - **Processing:** Normalize, PCA, scaling, feature completion and selection
- **Package 3: Model design**
 - **Model selection:** Implement different models, compare approach success
 - **Model architecture:** Hyperparameter search, validation
 - **Testing:** Separate model tests
- **Package 4: User webinterface**
 - **Visualization:** Choose framework, select design, implement
 - **User interaction:** Implement interactive interface, determine interaction limits
- **Package 5: Video creation**
 - **Planning:** Collect ideas, write storyboard
 - **Production:** Capture, cut, sound design
- **Package 6: Project management**
 - **Communication:** Team coordination, document milestones
 - **Organization:** Manage schedule, progress checking, weekly meetings

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

3 Workload distribution

We judge data collection to be one of the largest working packages. However, without data, most of the other packages won't be able to progress far. 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	
Energy consumption	Martin Schuck	Aladin Djuhera	Henrique 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	Software architecture, Video, Web interface
Felix Montnacher	Data collection, Data preprocessing, Web interface
Aron Endres	Data preprocessing, Model design, Video
Maximilian Putz	Model design, Data collection, Web interface
Henrique Soares Frutuoso	Data preprocessing, Model design, Web interface

Changes to the planned schedule might occur and resources will be redistributed accordingly.

4 Time Table

Naturally, data collection stands at the very beginning of the agenda. This is also required for the second milestone. Apart from defining interfaces, and abstract processes in each task, data preprocessing can only start once the data sets are acquired. The same is true for model design which relies on preprocessed data in order to benchmark performance. Web development can start independently from the model or the existence of sufficient data, and will therefore run in parallel with the previous work packages. Due to the sequential timeline of the first three work packages, we judge these work packages to be the most time-consuming tasks. Beginning video creation only makes sense at the end of the project. Proper management 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

The biggest risks to the project are potentially insufficient data on the Corona pandemic and data center/backbone electricity consumption. In case data sets show a poor quality or missing essential data for certain countries, the question may need to be limited or modified. This can be done by, for example, limiting the project to industrialized western countries where reliable data is likely to be more readily available. Furthermore, the pandemic produces highly atypical data with a very small sample size. This might lead to 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