To what extent will the Covid19 pandemic contribute towards reaching goals stated in the Paris Climate Agreement for Germany?

Data Collection Pipeline - 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 28, 2020

Submission structure

Uploaded on moodle for this submission:

- This main submission document, which gives an general overview over the project goal related to the data sources we decided to further move on with.
- PDF print of Jupyter notebook, that creates a central database (JSON) out of all individual data sources. Also it justifys the chosen and visualizes it.
- JSON File which comprises of all data that we will use in the next milestone (represents interface to milestone 3).

Provided in LDV GitLab repository:

- Jupyter notebook to create a central database, justify the chosen data and visualize it, located in the folder "src".
- Processed and raw data, located in the folder "data".
- Jupyter notebooks to, if needed, download data and process the raw data to standardized from, located in the folder "dataprocessing".

---

1 Project Goal

Paris Climate Agreement: In order to contribute to the objectives of the climate protection target decided in the Paris Climate Agreement, countries have submitted comprehensive national action plans to reduce their emissions. In a nutshell, these comprise the reduction of greenhouse gas emissions until the year 2030 by at least 40 % while taking greenhouse gas emissions of the year 1990 as reference for every participating country individually.

In short: In the project, we will use machine learning algorithms in order to predict and compare the greenhouse gas emissions in Germany with and without the impact of the COVID-19 pandemic. By comparing the two scenarios, we will be able to forecast the impact of the pandemic on Germany's climate targets on a long term basis and its effect on reaching the EU Climate Goals.

Approach: Our approach can be split into two scenarios. First, we want to predict the future greenhouse gas emission without the influence of COVID-19 pandemic on the basis of a given ground truth until 2018. In the second scenario, we will build a model that calculates the greenhouse gas emissions under the influence of the Corona crisis.
For the ground truth, we will look at monthly data of greenhouse gas emissions in order to later on capture the change during the pandemic more granularly and to make seasonal variation visible. Using our ground truth, we will use a regression model to predict the development of the greenhouse gas emissions without the influence of COVID-19 after 2018.
We will then use our ground truth on greenhouse gas emissions until 2018 to map key indicators that make up most of the greenhouse gas emissions or have a severe indirect effect on them. After having trained this model, we will use current data to measure the impact of the COVID-19 pandemic on greenhouse gas emissions for this year. Additionally, we will try different machine learning approaches to model the change in emissions during the pandemic and try to predict the future development in order to give a better estimate of the effect on reaching the EU Climate Goals.

Discussion: Using the measured difference between the two scenarios proposed in our approach, we will provide information on how the COVID-19 pandemic has already had an impact on our total greenhouse gas emissions in Germany. By additionally trying to give an estimate of the future development of the emissions we can further evaluate the impact of the crisis. We will scale our findings from Germany to all of the EU in order to classify how the direct and indirect effects of the pandemic has had an impact on reaching the EU Climate Goals in comparison to a scenario without the crisis.

2 Data

The indicators that we are mapping to the greenhouse gas emissions can be split into three main sectors, namely mobility, economy and the sector energy and households. For each sector we derive a machine learning model. The indicators that we use are presented in the following overview table. Additionally the data source for the ground truth, i.e. the greenhouse gas emissions for each sector is included.

Plots and an assessment if we expect the data to be relevant are provided in the Jupyter notebook.

Sector	Sub-Categories	Indicator	Sources	Characteristics
Mobility	Road traffic	Mobility Trend Google M1	https://www.google.com/covid19/mobility/	02/2020 - 06/2020, daily, processed to monthly, numeric Format: CSV
		Mobility Trend Corona Apple M2	https://www.apple.com/covid19/mobility/	01/2020 - 06/2020, daily, processed to monthly, numeric Format: CSV
		Mobility Trend Corona BAST M3	https://www.bast.de/BASt_2017/DE/Statistik/Verkehrsdaten/Verkehrsbarometer.html?nn=1820340	03/2020 - 06/2020, daily, processed to monthly, numeric Format: CSV
		Traffic Indices (like mobility trend) TomTom M4	https://www.tomtom.com/products/historical-traffic-stats/	01/2020 - 06/2020, daily, processed to monthly, numeric Format: CSV
		Traffic count Germany M5	https://www.bast.de/BASt_2017/DE/Verkehrstechnik/Fachthemen/v2-verkehrszaehlung/zaehl_node.html	2003 - 2018, hourly, processed to monthly, numeric Format: CSV
		Traffic count Bavaria M6	https://www.baysis.bayern.de/web/content/verkehrsdaten/dauerzaehlstellen.aspx	02/2017 - 04/2020, monthly, numeric Format: CSV, xls
		Driven km of cars M7	https://www.kba.de/DE/Statistik/Kraftverkehr/VerkehrKilometer/verkehr_in_kilometern_node.html	2014 - 2018, yearly, numeric Format: CSV, XLSX
	Railway and Bus traffic	Freight transport by railway M8	https://www-genesis.destatis.de/genesis//online?operation=table&code=46131-0004&bypass=true&levelindex=0&levelid=1592904937923#abreadcrumb	01/2005-02/2020, monthly, numeric Format: CSV, XLSX
		Passenger traffic (all) M9	https://www-genesis.destatis.de/genesis//online?operation=table&code=46100-0005&bypass=true&levelindex=0&levelid=1592905664759#abreadcrumb	2004-2019, per quarter, numeric Format: CSV, XLSX
	Shipping	Inland waterway transport (“Binnenschifffahrt”) M10	https://www-genesis.destatis.de/geness//online?operation=table&code=46321-0002&bypass=true&levelindex=0&levelid=1592906073196#abreadcrumb	01/1991-02/2020, monthly, numeric Format: CSV, XLSX
		Seafaring (freight transport) M11	https://www-genesis.destatis.de/genesis//online?operation=table&code=4633-0004&bypass=true&levelindex=0&levelid=1592906350275#abreadcrumb	01/2011-02/2020, monthly, numeric Format: CSV, XLSX
	Aviation	Aviation traffic statistic in germany M12	https://www-genesis.destatis.de/genesis//online?operation=table&code=46421-0012&bypass=true&levelindex=0&levelid=1592905811215#abreadcrumb	01/2011-04/2020, monthly, numeric Format: CSV, XLSX
		Scheduled lights Germany M13	https://www.oag.com/coronavirus-airline-schedules-data	01/2019 - 06/2019, 01/2020 - 06/2020, weekly, processed to monthly, numeric

				Format: XLSX
Economy	Economic Indicator	Consumer Price Index E1	https://www-genesis.destatis.de/genesis//online?operation=table&code=61111-0002&bypass=true&levelindex=0&levelid=1593074119569#abreadcrumb	01/1991-05/2020, monthly, numeric Format: CSV
		Number of Unemployment E2	https://www-genesis.destatis.de/genesis//online?operation=table&code=13211-0002&bypass=true&levelindex=0&levelid=1593362437069#abreadcrumb	01/2005-05/2020, monthly, numeric Format: CSV
	Stock Market	DAX E3	https://finance.yahoo.com/quote/%5EGDAXI/history?period1=631152000&period2=1592870400&interval=1d&filter=history&frequency=1d	01/1990-today, daily, processed to monthly values, numeric Format: CSV
		MDAX E4	https://de.finance.yahoo.com/quote/%5EMDAXI/history?period1=825552000&period2=1592870400&interval=1d&filter=history&frequency=1d	01/1996-today, daily, processed to monthly values, numeric Format: CSV
	Economy Sectors	Turnover Wholesale E5	https://www-genesis.destatis.de/genesis//online?operation=table&code=45211-0005&bypass=true&levelindex=0&levelid=1593073385190#abreadcrumb	01/1994-03/2020, monthly, numeric Format: CSV
		Turnover Retail Industry E6	https://www-genesis.destatis.de/genesis//online?operation=table&code=45212-0004&bypass=true&levelindex=0&levelid=1593074270719#abreadcrumb	01/1994-04/2020, monthly, numeric Format: CSV
		Turnover Hospitality Industry E7	https://www-genesis.destatis.de/genesis//online?operation=table&code=45213-0005&bypass=true&levelindex=1&levelid=1593076102258#abreadcrumb	01/1994-04/ 2020, monthly, numeric Format: CSV
Energy and households	Electricity generation	Carbon intensity of each type of power plant EH1	https://www.ipcc.ch/site/assets/uploads/2018/02/ipcc_wg3_ar5_annex-iii.pdf#page=7 Page 1335	Conversion factors, numeric Format: PDF
		Electric Energy Data of Germany starting from 2015 ordered by production type EH2	https://www.smard.de/home/downloadcenter/download_marktdaten/726#!?downloadAttributes=%7B%22selectedCategory%22:1,%22selectedSubCategory%22:1,%22selectedRegion%22:%22DE%22,%22from%22:1559340000000,%22to%22:1622584799999,%22selectedFileType%22:%22CSV%22%7D	01/2015-06/2020, 15min steps, numeric Format: CSV, XLS, XML
		Electricity generation, net heat generation, fuel input - Germany EH3	https://www-genesis.destatis.de/genesis//online?operation=table&code=43311-0002&bypass=true&levelindex=0&levelid=1592908210645#abreadcrumb GENESIS-Tabelle: 43311-0002	01/2002-03/2020, monthly, numeric Format: CSV
		Electricity balance of Germany EH4	https://ag-energiebilanzen.de/7-0-Bilanzen-1990-2017.html	1990-2001, yearly Format: XLS

		Electricity generation, net heat generation, fuel input - Germany EH5	https://www-genesis.destatis.de/genesis//online?operation=table&code=43311-0002&bypass=true&levelindex=0&levelid=1592908210645#abreadcrumb GENESIS-Tabelle: 43311-0002	01/2002-03/2020, monthly, numeric Format: CSV
	Weather	Cloudage, air temperature, amount of rainfall, sunshine duration EH6	https://opendata.dwd.de/climate_environment/CDC/regional_averages_DE/monthly/	01/1881 - 01/2020, monthly, numeric Format: TXT
	Households	Heating oil price EH7	https://www.boerse-online.de/rohstoffe/historisch/heizoelpreis/usd/22.5.2006_22.6.2020	22/05/2006-22/06/2020, daily, processed to monthly, numeric Format: XLSX
		Inflation rate EH8	https://de.statista.com/statistik/daten/studie/4917/umfrage/inflationsrate-in-deutschland-seit-1948/	1950-2019, yearly, numeric Format: XLSX
Greenhouse gas emissions	Emissions of Germany	Target values of greenhouse gas emissions of Germany GHG1	https://www.oeko.de/fileadmin/oekodoc/Sektorale-Abgrenzung-Treibhausgasemissionen-Datenbasis-20191217.xlsx Contact person for target values at “Umweltbundesamt”: Patrick Gniffke Tel: +49 (0)340 2103 2757 Patrick.Gniffke@uba.de	1990 - 2030, yearly, numeric 1990 - 2017: actually published data of “Umweltbundesamt” 2018 - 2030: currently estimated prognosis of “Umweltbundesamt” by Patrick Gniffke Format: XLSX

Table 1: Summary of Data Sources and Content

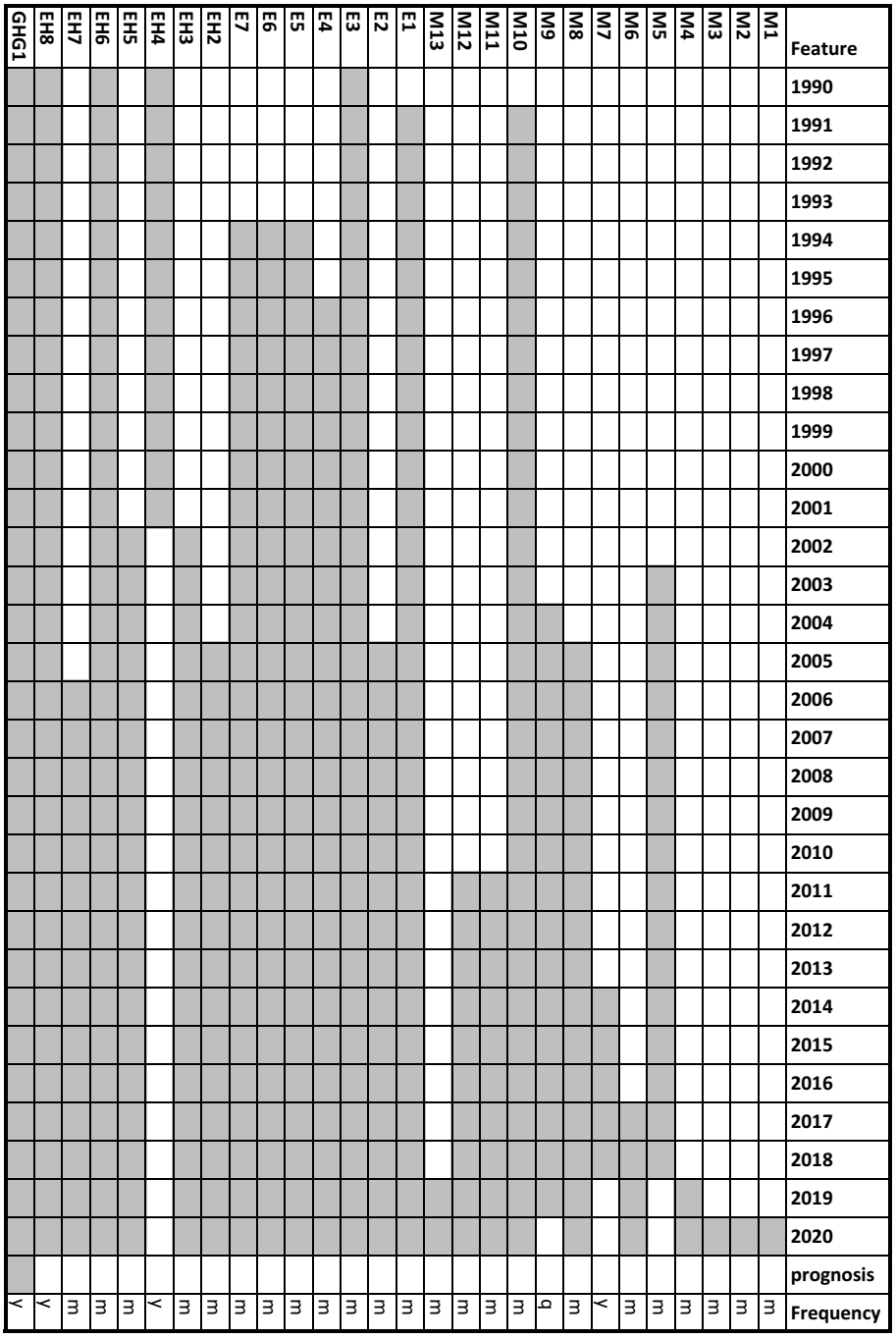

Figure 1: Coverage of the downloaded data