

Draft von Jochen Keilholz

Ziel ist die Modellierung und Vorhersage des 7-Tage Inzidenzwerts in Nürnberg

Inhalt:

1. Grundlagen
2. Datenquellen
3. Datenanalyse/Vorverarbeitung
4. Modelbildung
5. Visualisierung
6. Bereitstellung

1. Grundlagen

Die Idee ist die Einflussfaktoren/Features (Zeitreihen) des 7-Tage Inzidenzwerts (im folgenden 7TIW genannt) zu erforschen und auf Grundlage dieser den 7TIW zu modellieren. Der 7TIW wird als Target bezeichnet.

Folgende Features sollen in das Model einfließen:

- Wetter
- Kalenderspezifische Informationen (Feiertage, etc)
- Restriktionen der Regierung
- Mobilitätsbericht
- Werte angrenzender Lankreise
- Anzahl der geimpften Menschen in Bayern
- Hotspots/deren Entwicklung in NBG
- andere Coronawerte ?

Entwickelt wird in Python. Verwendete Daten werden aus open source quellen genutzt.

Ausgeliefert kann die Pipeline in verschiedenen Formen z.B. docker Container, streamlit app, python file...

2. Datenquellen

Die bisher recherchierten Datenquellen:

Für das Target: https://online-service2.nuernberg.de/aswn/ASW.exe?aw=HCOR_J03

Features:

Wetter: Daten vom DWD (open source) <https://pypi.org/project/dwdweather2/>

Kalenderdaten: <https://feiertage-api.de/>

Restriktionen der Regierung: <https://covidtracker.bsg.ox.ac.uk/>

Mobilitätsbericht: <https://www.google.com/covid19/mobility/>

Werte angrenzender Landkreise: tbd
Anzahl geimpfte Menschen in Bayern: tbd
Hotspots: tbd

3. Datenanalyse/Vorverarbeitung

Vorhersagebereich (alles und/oder):

- Mittelwert für die nächste/n Woche/n
- Trend, geht der 7TIW hoch oder runter
- Tageswerte für nächste Woche

Vorverarbeitung der Features:

Manche Zeitreihen werden nicht Tagesaktuell zur Verfügung stehen.

Für die "fehlenden" Werte sollen entweder lineare Interpolationen oder Schätzwerte durch andere Modelle verwendet werden.

Verarbeitung des Datums zu: Tag im Jahr, Wochentag, etc...

Evtl.: Scaling des Targets (Gleitendes Mittel über die 7 Tage) o.ä.

Wird sich zeigen wenn klar ist was vorhergesagt werden soll.

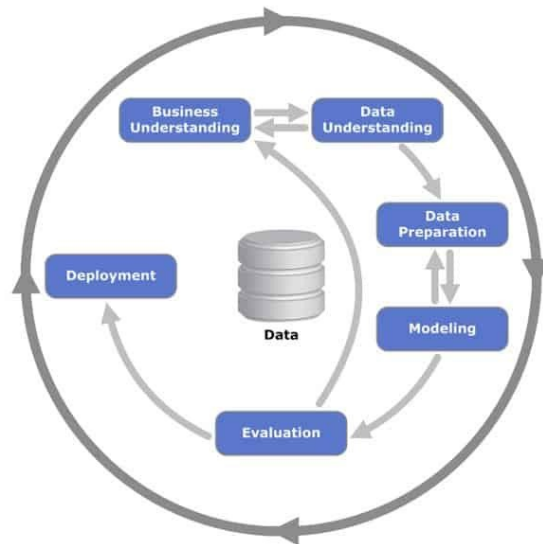
4. Modellbildung

Zwei vorschläge gibt es bis jetzt:

1. Lineares Model (fbprophet) einzeln oder mit zusätzlichen Regressoren. Mit diesem Model könnte man Vorhersagen treffen ohne täglichen eigenen Input
2. Gradient Boosted Trees (viele Entscheidungsbäume z.B. XGB). Mit dieser Art von Model könnte man auch mit den Inputparametern für die nächste Woche "spielen". Z.B. testen wie verhält sich der Output wenn es doch "viel" Wärmer würde. Mit diesem Model kann auch der Einfluss der einzelnen Features auf den Output geschätzt werden (wie wichtig erachtet das Model das Wetter, die Restriktionen, etc...)

5. Visualisierung

Pipeline wie crispedm:



Visualisierung der Outputs:

6. Bereitstellung

Je nach dem auf welcher Architektur der Prozess laufen soll:

- Docker container der die Outputs ablegt
- Jupyter Notebook
- Python script
- Streamlit im Browser, hier können auch die Inputparameter einfach verändert werden um zu testen wie die Vorhersage sich verhält!