

VORHERSAGE COVID-19 KENNZAHLEN

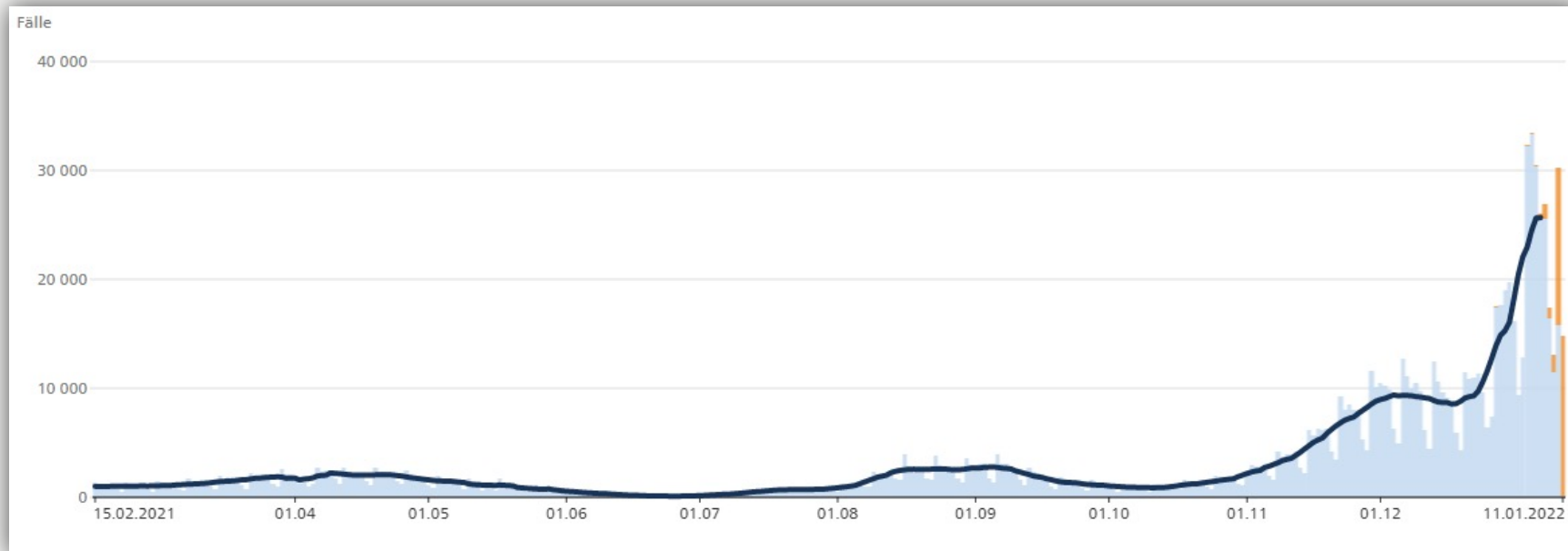
Anwendung und Vergleich von Machine Learning und Deep Learning Algorithmen

AGENDA

- Ziel des Projekts
- Methodik / Organisation
- Umsetzung
- Demo
- Resultate
- Diskussion der Resultate
- Erkenntnisse
- Fragen

ZIEL DES PROJEKTS

- Vergleich von verschiedenen Modellen zur Vorhersage der Covid-19 Fallzahlen
- Pro Modell Mean Squared Error (MSE), Dauer für Training und Prediction messen und vergleichen



METHODIK / ORGANISATION

- Code Verwaltungssystem: <https://github.com/maggi71/ml-corona>
- Entwicklungsumgebung: Visual Studio Code mit Anaconda Environment (Remote)
- Kommunikation: Wöchentliche Videokonferenzen
- Qualitätssicherung: Gegenseitige Code Reviews

UMSETZUNG

- Download der Daten von Opendata Swiss <https://opendata.swiss/de/dataset/covid-19-schweiz> als CSV Dateien
- Datenanalyse
- Aufbereitung und Bereinigung der Daten
- Implementierung der Modelle
 - Polynomiale Regression
 - Convolutional Neural Network (CNN)
 - Long short-term memory (LSTM)
- Optimierung der Hyperparameter mittels GridSearch und Cross Validation

HYPERPARAMETER

- Regression: Grad des Polynoms ($n=2, n=3, \dots, n=12$)
- CNN: `n_filter`, `n_kernel`, `n_epochs`, `n_batch`, `pool_size`, `activation_conv`, `activation_dense`, `optimizer`
- LSTM: (`n_layers`), `n_neurons`, `n_epochs`, `n_batch`, `optimizer`

DEMO

- Demo des LSTM Modells
 - Trainieren des Netzwerks
 - Prediction
 - Vergleich Prediction mit Ground Truth

RESULTATE

Algorithmus	Performance (MSE)	Laufzeit für Training	Laufzeit für Prediction
Polynomiale Regression	0.0266	0.0030s	0.0116s
CNN	0.0427	15.9702s	0.0747s
LSTM	0.0205	57.6358s	0.1137s

DISKUSSION DER RESULTATE

- Reihenfolge Performance
 1. LSTM
 2. Polynomiale Regression
 3. CNN
- Reihenfolge Laufzeit
 1. Polynomiale Regression
 2. CNN
 3. LSTM

Gute Datenqualität kostet Zeit

Rechenleistung für Training und Optimierung ist nicht zu unterschätzen

Python ist eine effiziente Programmiersprache für Data Science

Aufgrund des Datenschutzes waren Daten aus dem Geschäftsumfeld nicht möglich

Praxisbeispiel zeigt Komplexität von Machine Learning auf, im Vergleich zu konstruierten Beispielen

ERKENNTNISSE



FRAGEN

A decorative graphic on the left side of the slide consisting of three parallel, wavy vertical lines. The innermost line is yellow, the middle line is white, and the outermost line is a light beige color. They all extend from the top to the bottom of the frame.

Danke fürs Zuhören