

Dataset

Simultaneous Time Series Forecasting on the World's COVID-19 Daily Vaccinations

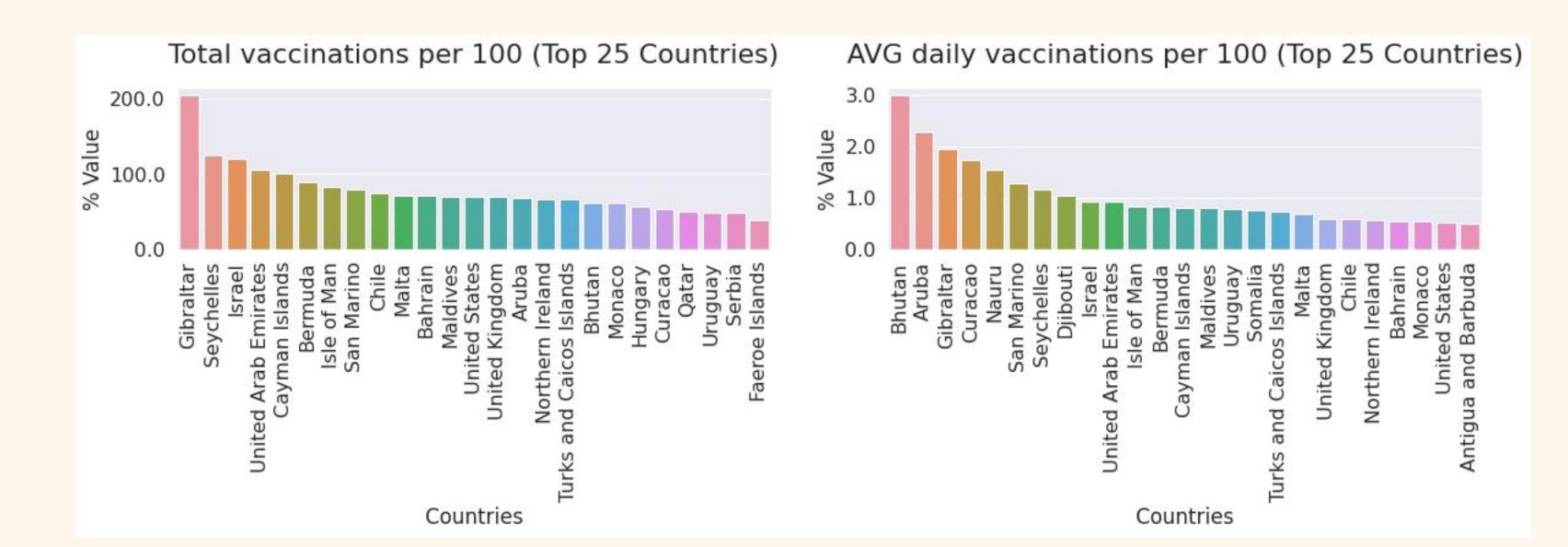
Konstantinos Georgiou

Course Advisor: Michela Taufer

The University of Tennessee, Knoxville

Motivation

- The distribution of COVID-19 vaccines affects the health of billions of people as well as the state of the world's economies
- Many efforts have been made to extract useful insights from these data, but most of them are comparative analysis between two or more countries
- As of today, no method attempted to simultaneously predict the number of daily vaccinations of all the countries by utilizing the correlations between them
- We introduce a method that uses Encoder-Decoder Long Short-Term Memory Networks With Multivariate Inputs and Walk-Forward Validation of 10 days.



Dataset

- Contains daily vaccination data for 193 different countries and 135 dates
- 14230 15-dimensional data from which 8 dimensions where used
- The dataset has many null values, most of which can be inferred from other values



Workflow

New Data Training Evaluation Preprocessing Dataset every day Group the Dataset in 10-day windows and Visualize: Group the Dataset in 10-day windows and **Columns kept:** Data columns (total 15 columns, 14230 Entries): discard the excess days Histograms with countries and dates that discard the excess days • iso_code **Non-Null Count** Column • 126 days → 12 10-day groups + 6 have the highest and lowest errors • 126 days → 12 10-day groups + 6 Date days discarded The test and predicted sets' daily daily_vaccinations days discarded 14230 non-null country Split in Train (~80%) and Test (~20%) set vaccinations per hundred timeseries for total_vaccinations Split in Train (~80%) and Test (~20%) set 14230 non-null iso_code people_vaccinated specific countries 14230 non-null people_fully_vaccinated 8486 non-null total vaccinations total_vaccinations_per_hundred 7850 non-null people_vaccinated people_vaccinated_per_hundred people_fully_vaccinated 5732 non-null people_fully_vaccinated_per_hundred daily vaccinations raw 7101 non-null **Pivot the Countries in the Dataset by** lstm_input: InputLayer daily_vaccinations_per_million daily vaccinations 14033 non-null Calculate RMSE: grouping rows by date 8486 non-null total_vaccinations_per_hundred The average of all the predictions New Columns (1409 Columns, 126 Entries): people_vaccinated_per_hundred 7850 non-null Average per Country Date people_fully_vaccinated_per_hundred 5732 non-null lstm: LSTM Average per Date daily_vaccinations_country_1 daily_vaccinations_per_million 14033 non-null Infer nulls for the non-percentage 14230 non-null vaccines float columns from their daily_vaccinations_country_N 14230 non-null source name relationships people_vaccinated_country 1 repeat_vector: RepeatVector 14230 non-null source_website Gather Predictions people_vaccinated_country_N Recreate the test set and the predictions lstm_1: LSTM in the original tabular format Recalculate the per hundred values Countries Join the two datasets metadata using the population of the and drop rows with

Normalize

Drop recent dates when

many countries are missing

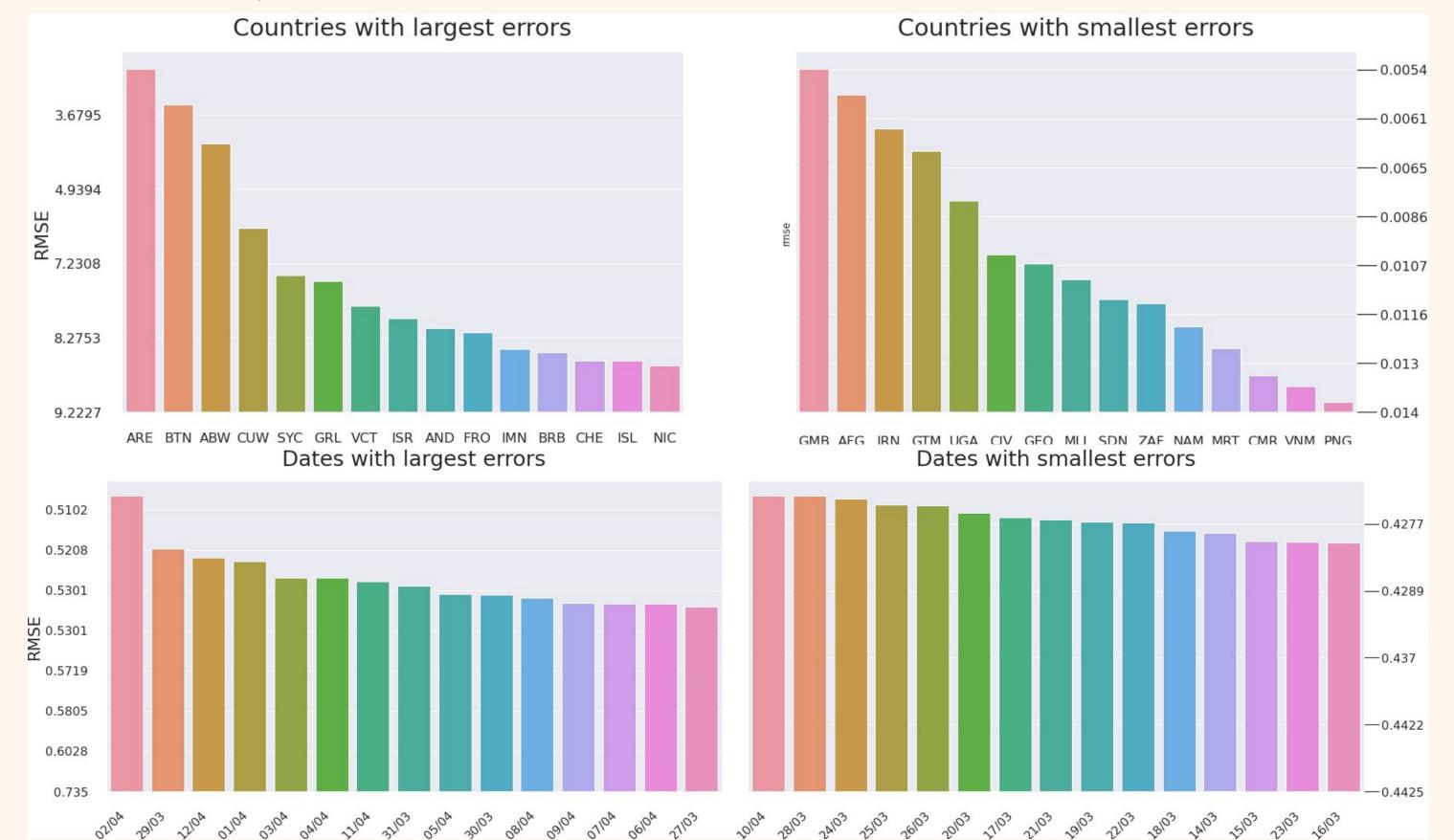
Evaluation

countries

Average RMSE: 0.246174

- Out of the 170 countries, 164 had mean RMSE less than 1.0, 153 less than 0.5, and 73 less than 0.1 Min: 0.00023, Max: 5.5133
- Out of the 30 dates, 25 had mean RMSE less than 0.30, 15 less than 0.25, and 11 less than 0.20 o Min: 0.16928, Max: 0.35180

missing countries



Conclusions

Results

Predict the Daily Vaccinations per

hundred people on the Test set

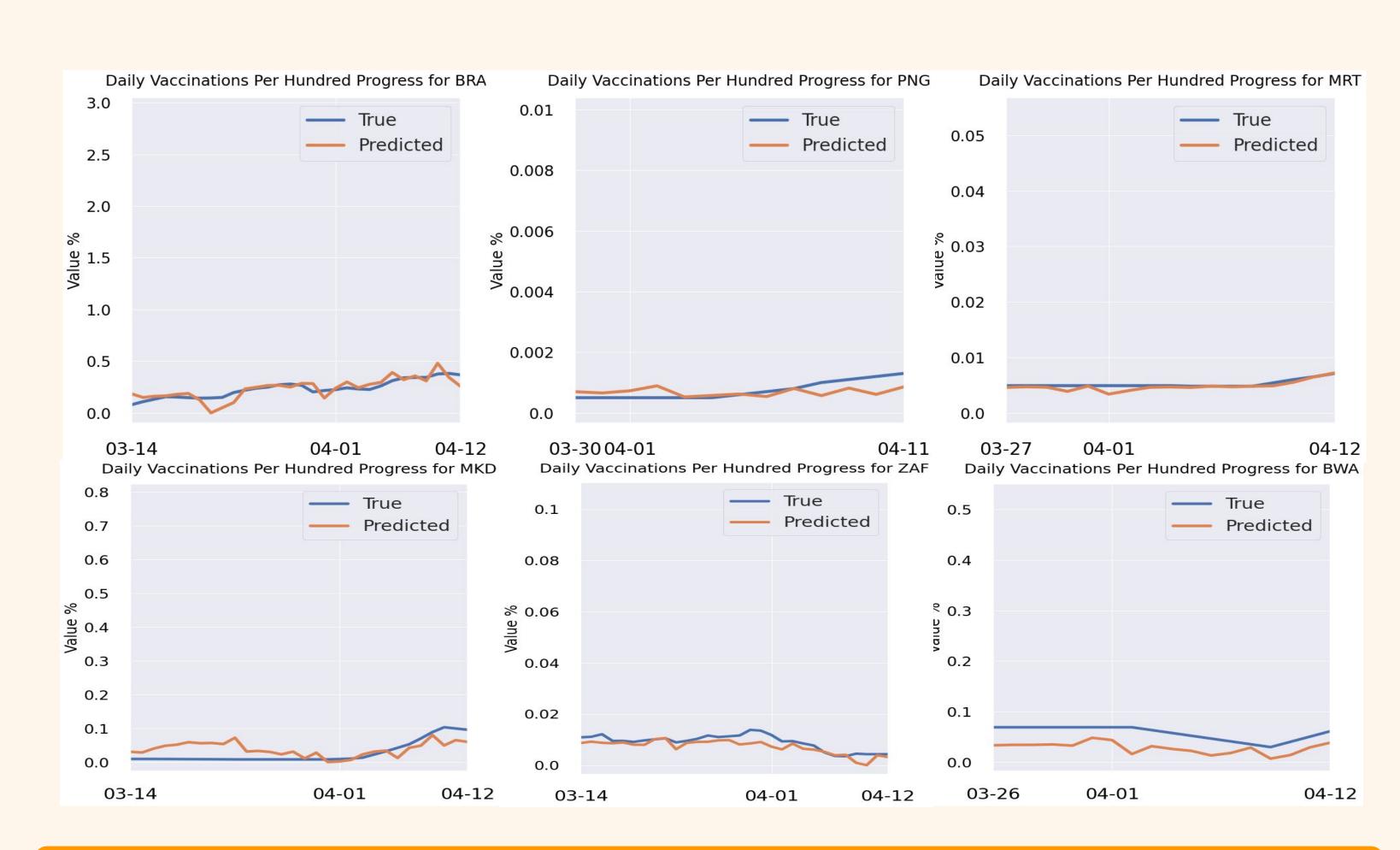
predict the 10 next

• Each time, use the 10 previous days to

time_distributed(dense): TimeDistributed(Dense)

time_distributed_1(dense_1): TimeDistributed(Dense)

True VS Predicted Daily Vaccinations per million for 6 countries



Future Work