Pandemic Forecasting

The National Institute of Infectious Diseases[1] has been tasked with tracking the COVID-19 pandemic in South Africa and wants to know from Isazi how many new cases there will be over the following week.

To get recent data you will need to scrape COVID-19 statistics for the country from the internet. Using this data you will then need to generate a forecast for the following seven days from when the forecast is run.

Your solution should consist of the following components:
1. Ingestion
   - Get the HTML source from a COVID-19 website such as worldometers[2].
   - Parse the webpage source to get the number of daily cases in the recent past.
   - Note: Any API queries that require sign-up/keys are not allowed for this task.
   - Note: You can assume the website structure won't change in the future.
2. Modelling
   - Use an applicable statistics/machine learning method (e.g. ARIMA, ETS, GLMs, neural nets, …) to make a forecast for the number of new cases over the following seven days.
   - Note: The model should make predictions starting from the day after when the model is run.
   - Note: The model should run 'in-real-time'. You therefore cannot use any downloaded data for your model.
3. Simple reporting
   - Write your results to a .csv file with the following column headings: `date` and `new_cases`. You should also write your forecast as a timeseries plot to a `.png` file.
   - The institute also wants advice about how their system could be improved in the future. In a .txt file, write a short paragraph about other kinds of data you would like to include to improve the forecast accuracy.

The above steps should run end-to-end with a single call to a script/notebook. Your solution must be written using only open source tools (e.g. golang, python, R, ...) so we can run your solution and check your results. If necessary, in a `README` file, specify any external libraries or packages that you used.

---

[1] A fictitious government institute
[2] `https://www.worldometers.info/coronavirus/country/south-africa/`

Bonus Task

The completion of this task is not required but it will help us assess some additional capabilities that are in line with the role.

The National Institute of Infectious Diseases would prefer to have a web interface where they can easily view the forecasting results, instead of relying on periodic updates via CSV files.

In order to achieve this, you would need to expose your forecasting results via a very simple web application. This would require the following:

- Making the forecasting results accessible via an API endpoint.
- Make a very simple frontend for the web application which displays the forecasting results obtained from your API - both in a table and as a timeseries plot.
  - You may use any plotting/graphics library e.g. *chartjs* as long as it does not require a license.
  - You may use any frontend framework (or none at all).

Note:
- This web application does not need to be hosted anywhere, simply provide us with the necessary files and instructions on how to run the API and frontend locally in the `README` file.
- You do not need to have any additional buttons, inputs, etc. on the frontend - you need only show the forecasting results.
- You do not need to take into account authentication and authorization at all.