#ref #ret

0.1 | Record of understanding:

The writing style here is more similar to my notes, which I personally use as a record of my
understanding. If you would like this to be written up more formally, please let me know!

0.1.1 | 1. Paper Title:

A COVID-19 pandemic AI-based system with deep learning forecasting and automatic statistical
data acquisition: Development and Implementation Study

0.1.2 | 2. What question or phenomenon was being investigated in this study and why?

- COVID is a large problem, almost all of the AI studies around it are region specific or centered around a single country
- CPAIS (COVID-19 Pandemic AI System) aims to solve this by automatically compiling databases to form a worldwide dateset
    - Including governmental responses
    - allows heatmap viz of different policies in different countries,
    - and predictive modeling, ofc.

0.1.3 | 3. What background information did you need to understand in order to understand the question, main experiment, main finding, and significance? (i.e. what did you find yourself digging into in the introduction or through outside sources; what was the most relevant background info?)

ARIMA: *Auto Regressive Integrated Moving Average*

- Type of Auto Regressive (AR, not Augmented Reality) model
    - used for processes which change over time (economics, web traffic, ect.)
- Very similar to ARMA, but with a more complex stochastic term
- Uses time series data
    - ie. use differences between values rather than the actual value
    - p, d, q
        - p: order of AR term, or time lag
        - d: "degree of differencing"
        - q: order of MA term, or size of moving average window

    Predicted Y_t = Constant + Linear combination Lags of Y (upto p lags) + Linear Combination of Lagged forecast errors (upto q lags) - Good article

MLP: *Multilayer Perceptron Neural Networks*

- Just a feed forward network, but with hidden layers. Pretty ambiguous term, left relatively ambiguous in the paper.
- Fancy name for a vanilla neural net
- They highlight: designed to solve non-linearly separable problems

    - with sigmoid

- They used nnfor, which means they used R.

MAE (vs RMSE):

- Mean Absolute Error, which I haven't used before, is like RMSE (root mean square error) but less punishing for bigger errors.

MAPE: *Mean Absolute Percentage Error*

- aka Mean Absolute Percentage Deviation
- loss function, used for measuring forecast accuracy.
- average of percentage errors
- generally not good..

0.1.4 |4. What was the main thing the researchers found out and how did they do so?…

1. what were the main (1-2) experiments?

    1. Dateset gen and renewal:
        1. using crawler on source
        2. integrating the data into existing database
        3. statistical analysis with predefined procedure
            1. The experimentation comes in at this step with different deep learning models and techniques
    2. **Main Models:**
        1. LSTM: *long term short term memory*
            1. A type of recurrent neural network which uses a 'cell state' and forget gates to have both long and short term memory
            2. better forecast accuracy than other models
        2. ARIMA: *Auto Regressive Integrated Moving Average*
            1. statistical learning model with time series regression

2. what data did they generate?

    1. Generated a worldwide automatically updating dateset for COVID-19 with their automation technology
    2. And also, generated consistent forecasting data of the next 14 days
    3. Finally, generated data on the correlation of government policies and COVID-19

3. what does that data mean?

    1. The dateset, innit of itself, means nothing.
    2. the forecasting data means what we can expect for the next two weeks
    3. the correlation data *can* mean what governmental policies work best for dealing with COVID-19

0.1.5 | **5. What was the significance or larger impact of the main finding?**

- the dateset itself is useful for the entire world doing data analysis
- the forecasting data can help with rapid policy changes and preparations. Being able to see two weeks into the future, especially at the beginning of the pandemic, was and is incredibly useful.
  - they also made ways to visualize the data, as human understanding is ultimately the most important at this point
- they didn't make very substantive claims about governmental policies
- instead, they mostly made a very useful tool for data viz, prediction, and comparison. In the actual experimentation with the types of models, as expected, LSTMs did better.
  - they were experimenting with ARIMA, but ultimately, LSTMs win again.

0.2 | **Reflection Questions**

1. **What paper did you choose and why did you choose it?**

   1.   A COVID-19 pandemic AI-based system with deep learning forecasting and automatic statistical data acquisition: Development and Implementation Study
   2. I was interested in doing some similar things earlier, so this paper caught my eye.

2. **How did you go about trying to understand the paper that you chose? What was your reading/understanding process like and why did you employ that strategy?**

   1. Generally speaking, my strategy was to get background on the topic (normally by reading the abstract) and then follow my curisity throughout
   2. I read the abstract of the paper, then learned about the terms I didn't understand from the Introduction and *Materials and Methods* section. I also used the handy cmd-f functionality to search through the document. However, a lot of my time was spent looking at other articles online for deeper explanations. I then jumped

3. **What did you find challenging about trying to understand your paper? Although the task may have felt generally challenging, try to get specific here.**

4. **What do you think you could try next time that might improve your process?**

5. **What type of previous experience do you have with reading papers from the scientific literature (either review articles or primary research)?**