Project COVID-19

By Maram Edrees

Project COVID-19

Outline

- Problem Statement
- Solution to the Problem: Project COVID-19
- Building the Chatbot
- Chatbot Performance
- Benchmarking
- Challenges
- Next steps

Problem Statement

There is an urgent need for a platform from where accurate information and updates about COVID-19 can be obtained.

Solution to the Problem: Project COVID-19

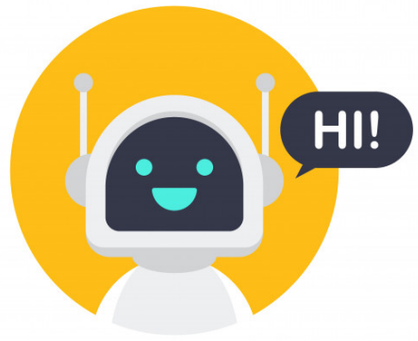
What is Project COVID-19?

- It is an initiative that builds an interactive COVID-19 chatbot which leverages AI and deep learning technologies.
- It uses information from official sources for chatbot training.

Solution to the Problem: Project COVID-19

Meet Bot, your COVID-19 chatbot!

- Facilitates access to reliable information regarding the COVID-19 pandemic
- Provides accurate and updated local and international COVID-19 figures
- Reports the latest news on COVID-19
- Fights misinformation



Bot updates to accommodate the daily, weekly, or hourly updates published by LiveCorona, WHO, MOH, or MOI.

Building the Chatbot

Step 1. Collecting the Data

Data Sources

REST API

Data Extraction
- Selenium
- Twitter APIs

Raw Data Parsing
- Remove tags
- Remove non-English text

Data Cleaning
- Remove spaces and symbols

*TrackCorona is a website sponsored by Google Maps and Wikipedia

Step 1. Collecting the Data

Clean Data:

Data Source	Data Rows	Data Features
TrackCorona Website	185	4
MOH & MOI Tweets	316	7
MOH FAQs Website	11	3

Sample Data:

Location	Confirmed_Cases	Recovered_Cases	Active_Cases	Fatalities
United States	598340	44364	529189	24787
Spain	172655	67504	87001	18150

Created_at	Source	Tweets
2020-04-10 18:15:00	Ministry of Health	Know the difference between medical quarantine...
2020-04-10 17:24:29	Ministry of Health	Responsible measures be taken before, during a...

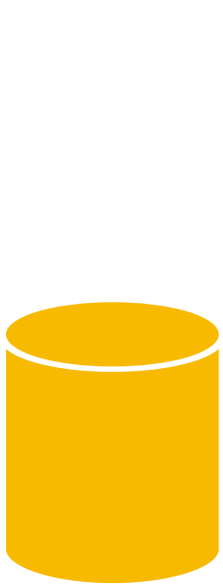
Questions	Answers
What are Coronaviruses?	Coronaviruses (CoV) are a large family of viru...
What are the species of coronaviruses that hav...	The SARS-CoV was transmitted from civet cats t...

Step 1. Importing Libraries

The libraries used were:

1. NLTK (natural language processing library)
2. Tweepy (Python library for access Twitter API)
3. Keras (neural network library)

Step 2. Preprocessing the Data



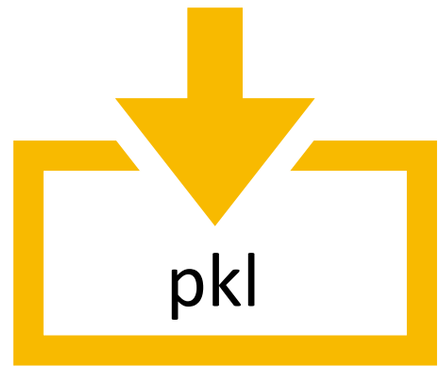
Preprocessing Text

- Tokenization: breaking sentences into words
- Stemming: reducing inflectional forms of words to a common root

Example: *healthcare*, *healthier*, and *healthy* will be replaced with *health*

Feature Engineering

- Intents: contain a collection of tags with their corresponding patterns and responses
- Classes and documents: contain intents associated with patterns



Step 3. Creating the Training Data

1. Converting input patterns into numbers:
    - Creating a list of zeroes equal to the total number of words
    - Setting a value of 1 only to indexes that contain the word in the patterns
2. Creating the output by setting 1 to the class input to which the pattern belongs

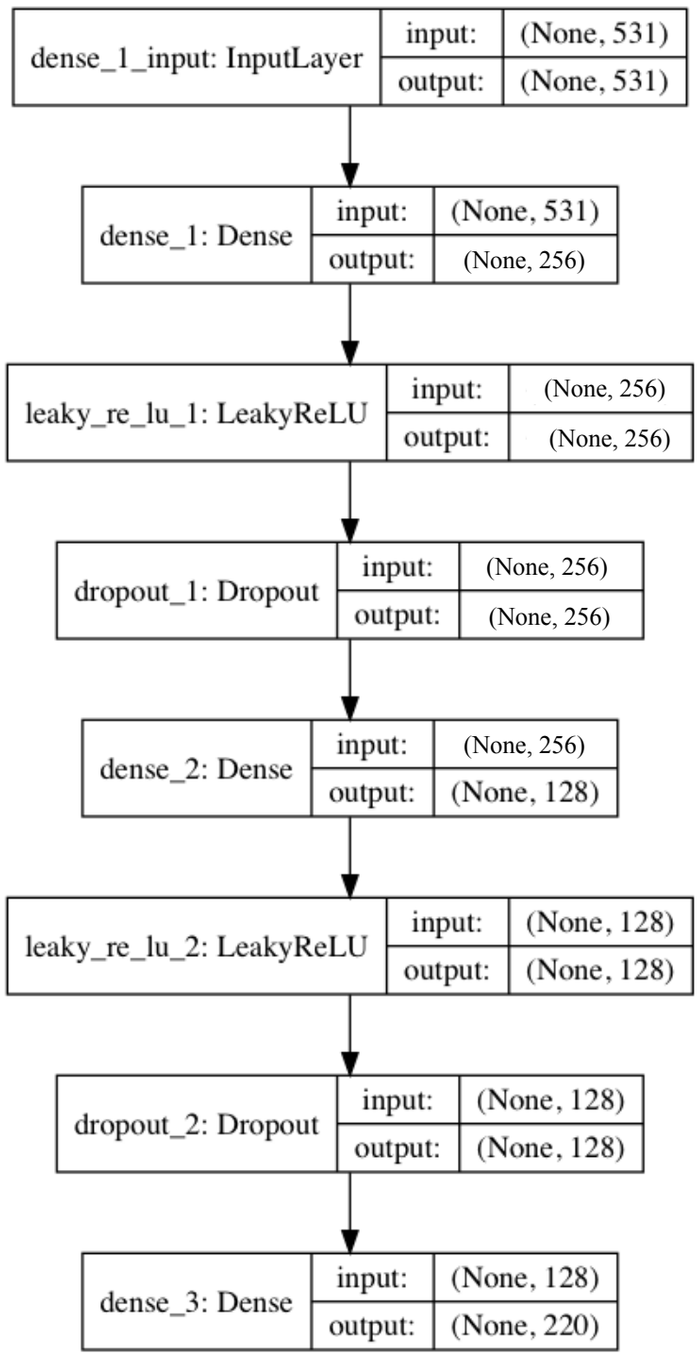
Step 4. Training the Model

Deep Neural Network — Sequential Model

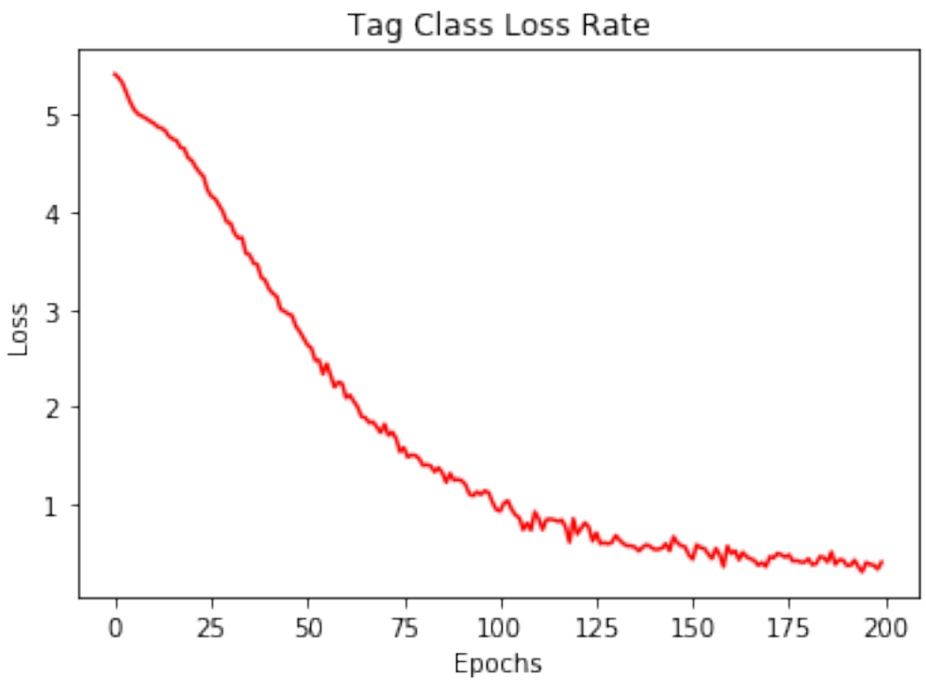
1. The sequential neural network consists of three dense layers:
    - First layer (256 neurons)
    - Hidden layer (128 neurons)
    - Last layer (the number of neurons is equal to the number of classes)
2. Overfitting was reduced by adding dropout layers
3. Stochastic gradient descent (SGD) optimizer and data fitting were used to train the model
    - SGD is a parameter tuning technique used to determine the most optimal model
    - SGD randomly selects a parameter update for each iteration
4. The Nestrov accelerated gradient was employed
5. After being trained for 200 epochs, the model was saved

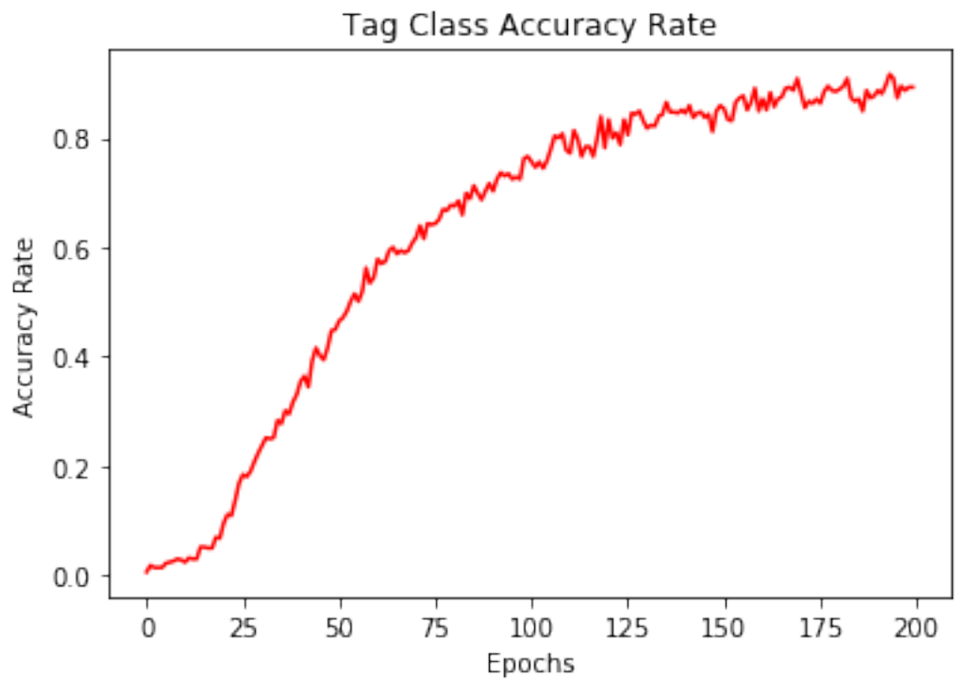
Step 4. Training the Model

DNN Layers:



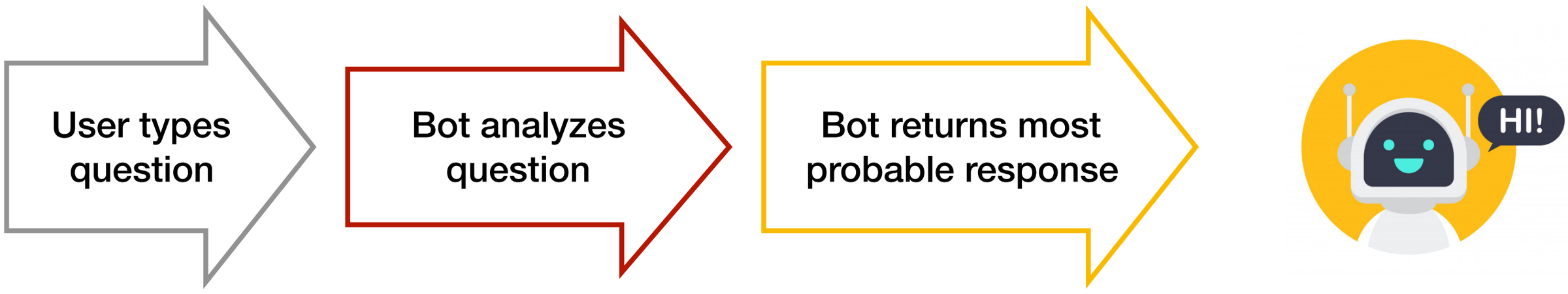
Step 4. Model Evaluation





Step 5. Interacting with the Chatbot

1. A interactive graphical user interface was created for the chatbot.
    - Tkinter module was used to build the structure of the desktop application.
2. The bot was programmed to predict the class to which a pattern belongs in order to retrieve an answer that matches a query.



Model Performance

	Model Variations	Correct Answers	Incorrect Answers	No Answer
1	Baseline model Relu: 3-layer model 1st layer (128 neurons) 2nd layer (64 neurons)	2	3	-
2	Leaky Relu: 3-layer model 1st layer (128 neurons) 2nd layer (64 neurons)	2	3	-
3	Leaky Relu: 3-layer model 1st layer (256 neurons) 2nd layer (128 neurons)	4	1	-
4	Leaky Relu: 4-layer model 1st layer (256 neurons) 2nd layer (128 neurons) 3rd layer (128 neurons)	2	3	-
5	Leaky Relu: 5-layer model 1st layer (256 neurons) 2nd layer (128 neurons) 3rd layer (128 neurons) 4th layer (128 neurons)	1	-	4
6	Leaky Relu: 4-layer model L1 regularization (Lasso) 1st layer (256 neurons) 2nd layer (128 neurons) 3rd layer (128 neurons)	-	-	5
7	Leaky Relu: 4-layer model L2 regularization (Ridge) 1st layer (256 neurons) 2nd layer (128 neurons) 3rd layer (128 neurons)	-	-	5

Bot Performance

Chatbot metrics are in their early stages of development — few references for consultation are available.

Interactions per User

Measuring the interactions between users and Bot

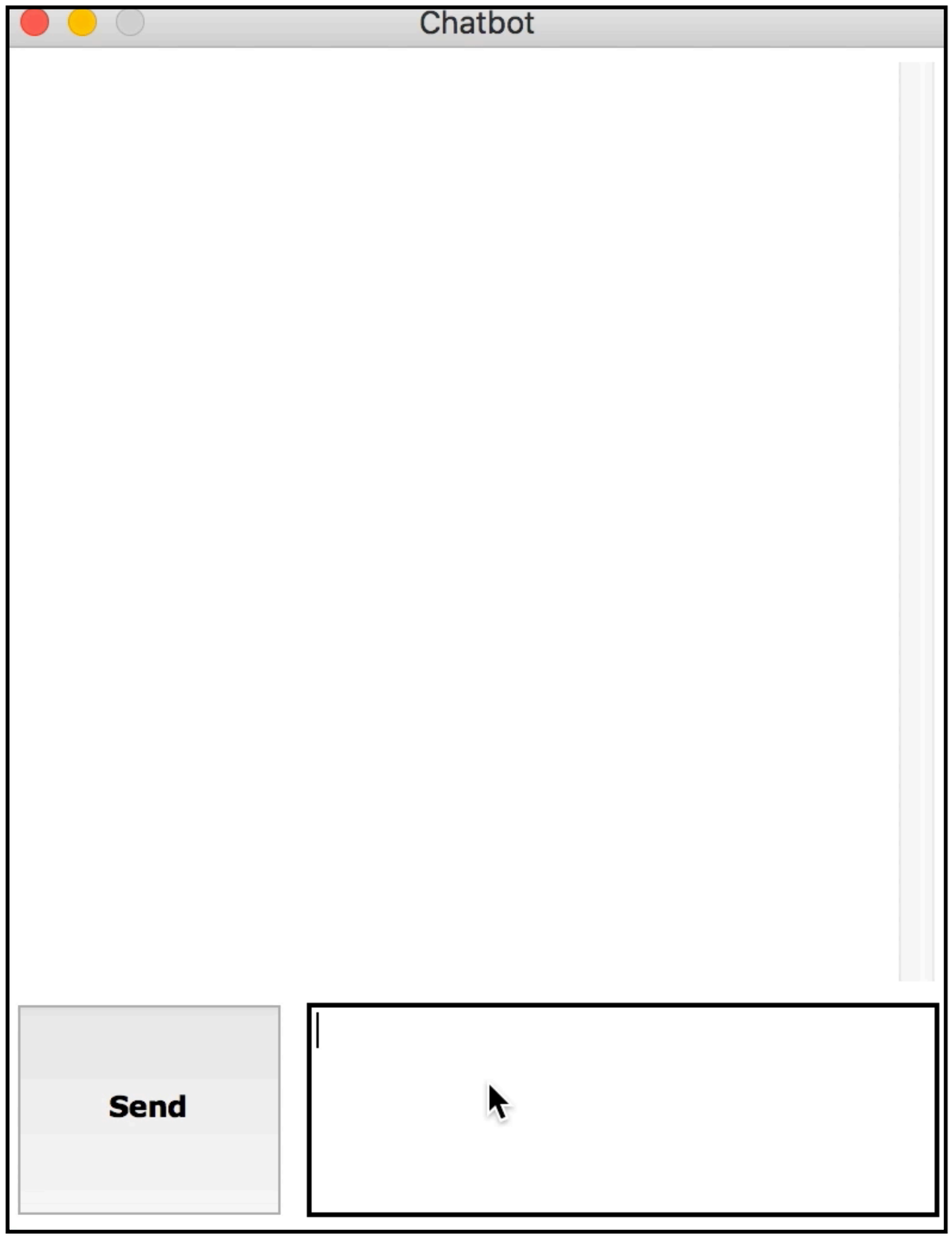
- Bot was tested several times by different users.
- Several technical errors were noted during the testing, and these errors have already been addressed.

Retention and Goal Completion

Repeat users and usage purposes

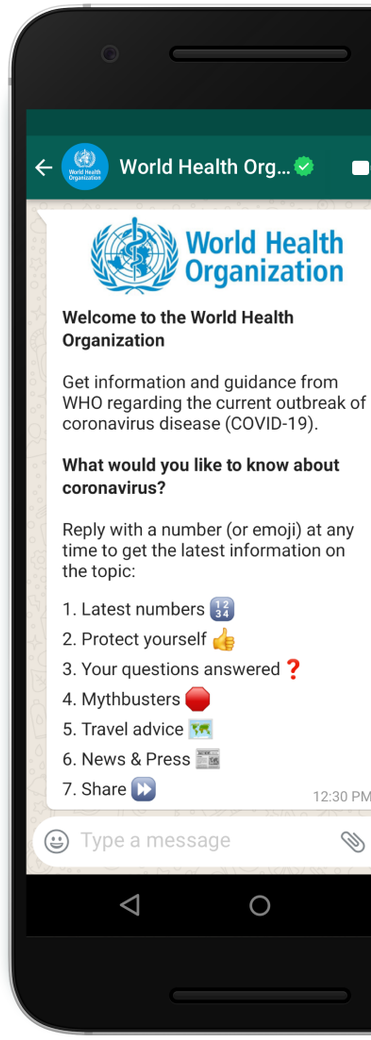
- Most users who tested Bot would engage with it daily (number of users: 2−6)
- Motivation behind using Bot is checking the latest COVID-19 information from MOH and MOI.

Model Performance Demo



Chatbot Benchmarking

Characteristics	BOT	World Health Organization Chatbot
Conversational AIs	Yes	No
Country-specific figures	Yes	No
Local information	Yes	No
Region-specific figures	No	Yes
Social media platform	No	Yes



Challenges

- Tuning the parameters (trial and error)
- Operating Bot on Jupyter Notebook is inconvenient
- Hourly updates must be monitored daily so that changes in the format of data sources (websites) are incorporated during data parsing.

Next Steps: Future Work

- Deploy the model on a social media platform
- Add more features to the chatbot
- Include Arabic responses

References:

- Laetsch, T., 2020. Web Scraping With Python, accessed on 19 Mar, 2020:(https://campus.datacamp.com/courses/web-scraping-with-python/introduction-to-html?ex=1).
- Navlani, A., 2019. Text Analytics for Beginners using NLTK, accessed on 19 Mar, 2020:(https://www.datacamp.com/community/tutorials/text-analytics-beginners-nltk).
- Neff, A., 2019. METRICS UNCOVER AND TRACK THE VALUE OF CHATBOTS WITHIN CUSTOMER SERVICE, accessed on 19 Mar, 2020:(https://www.icmi.com/resources/2019/metrics-uncover-the-value-of-chatbots-within-customer-service).
- Jassova, B., 2019. Key Chatbot KPIs: How to Measure Chatbot Performance?, accessed on 19 Mar, 2020:(https://landbot.io/blog/chatbot-metrics-kpi/).

Thank you for listening.

شكرا على حسن الاستماع