Prediction of COVID-19 Cases in Ontario

Ali Sedigh

September 30, 2020

Agenda

Background & Objectives

Data Description

Exploratory Analysis

Data Preparation

Method

Assumptions

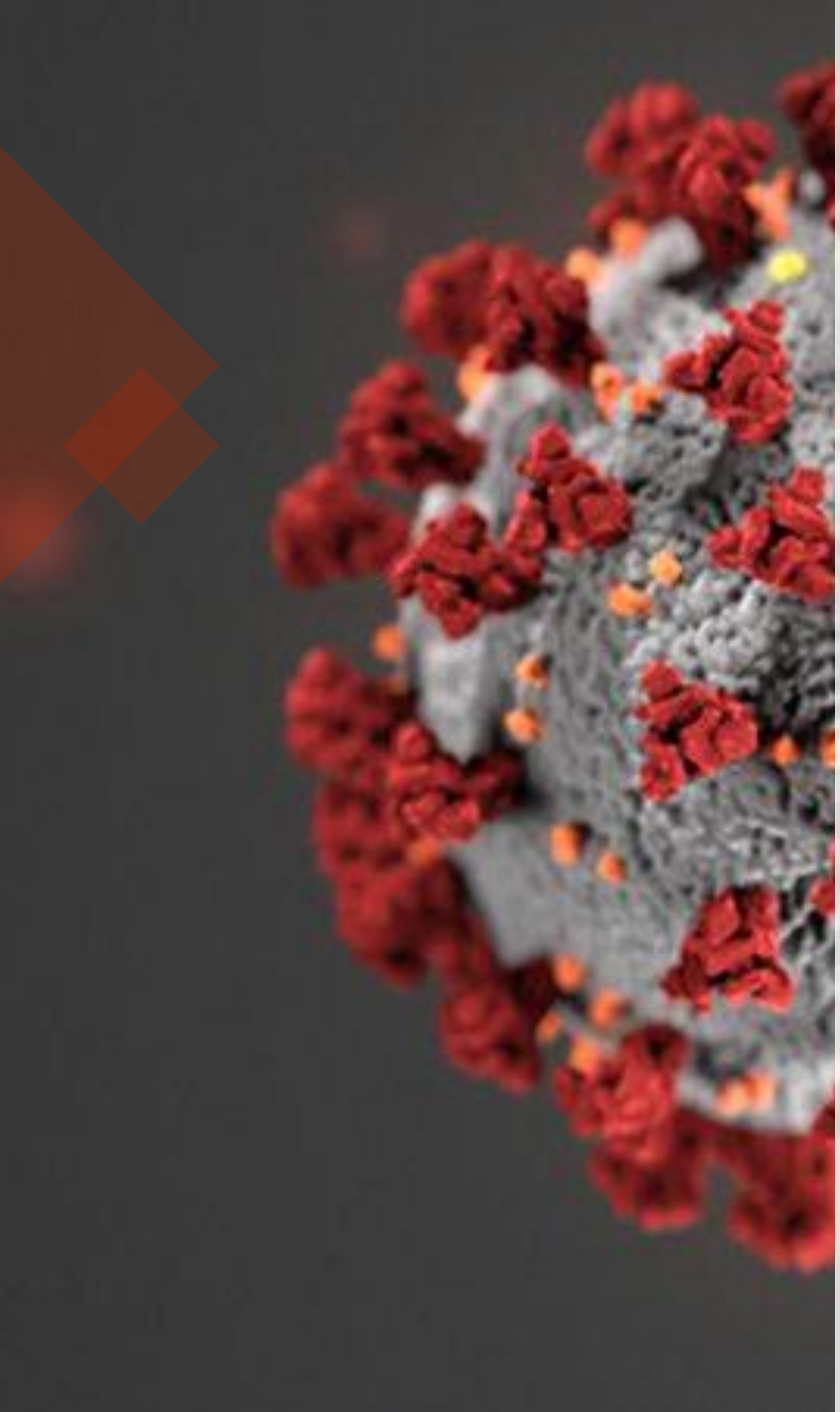
Accuracy Test

Results

Limitations

Next Steps

Appendices



Background & Objective

- The outbreak of coronavirus disease 2019 (Covid-19) has exposed the world to great challenges and is a serious concern for public health. The outbreak started in Wuhan, China, in December 2019 [1], [2].

- There is a lack of information and uncertainty about this outbreak, making it important to understand its dynamic behavior. Forecasting the outbreaks behavior over time can provide useful insights into the epidemiological situation [3],[4].

- **Objective:** To create deep learning models to predict weekly COVID-19 cases based on the past number of confirmed cases in Ontario, Canada.

Data Description

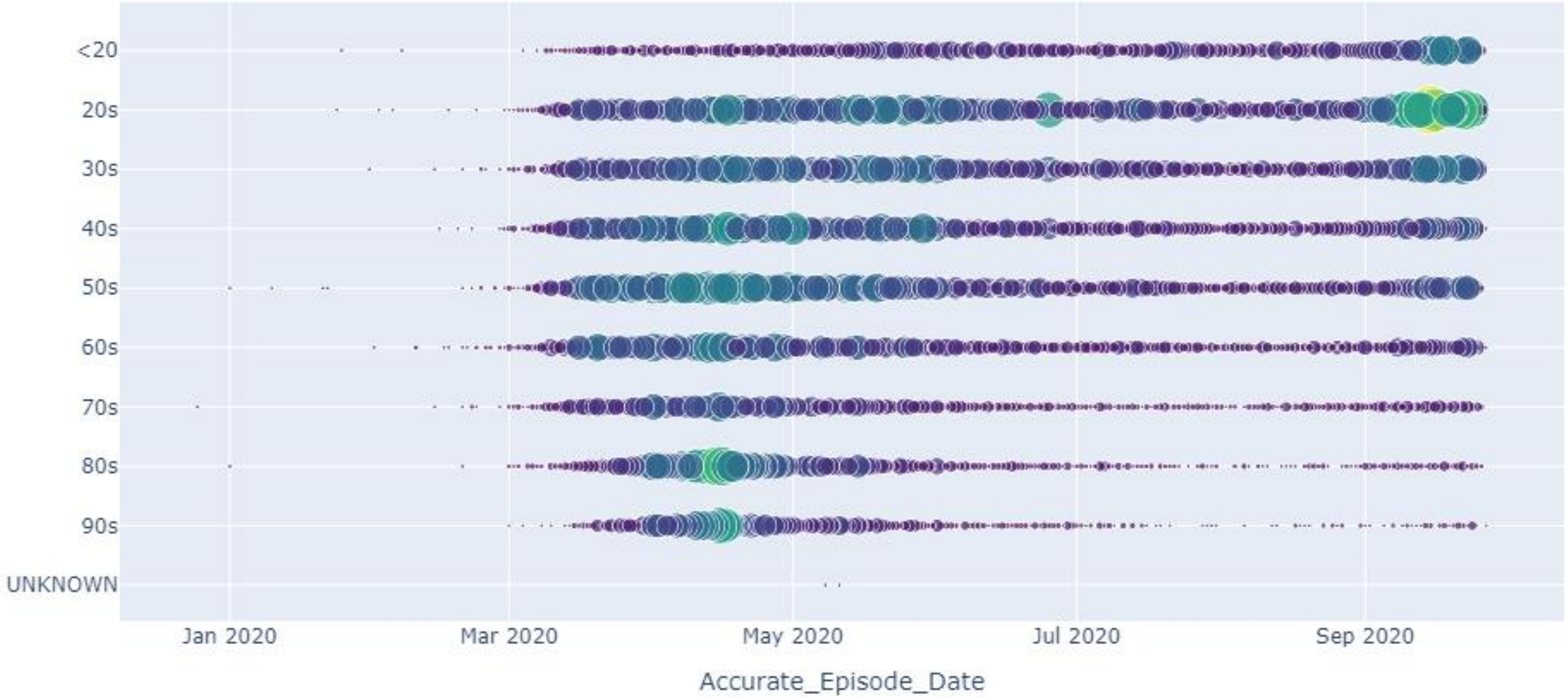
- Confirmed positive cases of COVID-19 in Ontario: The dataset compiles daily snapshots of publicly reported data on 2019 Novel Coronavirus (COVID-19) testing in Ontario as of September 28, 2020. https://data.ontario.ca/dataset/confirmed-positive-cases-of-covid-19-in-ontario

- Statistics Canada, 2016 Census of Population[5].

- Public Health Ontario, Ontario COVID-19 Data Tool[6].

Exploratory Analysis

- Confirmed Cases Over Time by Age Group



Exploratory Analysis

Observations

Time Series Data

- Looking into history of previous days is of great importance.

Age

- Lower number of new cases reported for adults over 60. Also adults over 60 years old are reported by Government of Canada to be at risk of more severe disease or outcomes [7].

Public Health Units

- Average confirmed COVID cases per day in Greater Toronto Area (GTA) Public Health Units (Toronto, York Region, Peel, Durham Region, Halton Region) is among the top 9 of all public health units [Appendix-2].

Data Preparation

Two PHU groupings: GTA and Other

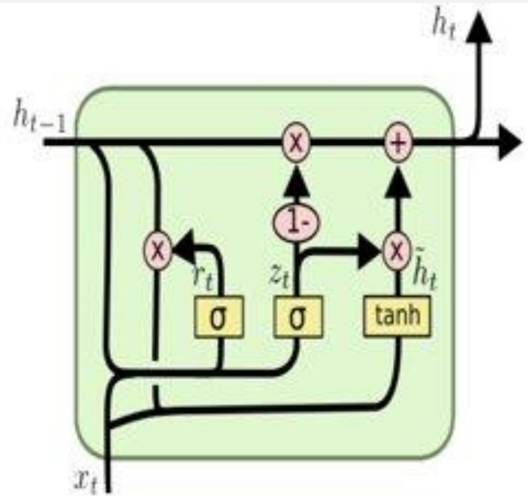
Two age categories: <60 and >=60

Group	Estimated Population [5][6]
Ontario	14,864,428
GTA PHUs and age <60	5,615,653
GTA PHUs and age >=60	1,667,943
Other PHUs and age <60	5,844,821
Other PHUs and age >=60	1,736,011

Note: In 2016 Census in Ontario, 77% of population were under 60 years of age and 23% were equal or over 60 years of age[5]. It is assumed that the same distribution is still true in 2020 and the distribution is even throughout the province (i.e. PHUs).

Method: Long Short-Term Memory (LSTM)

- A Recurrent Neural Network (RNN).
- Designed for Sequence Prediction problems and time-series forecasting nicely fits into the same class of problems.
- Many to one LSTM model is used.



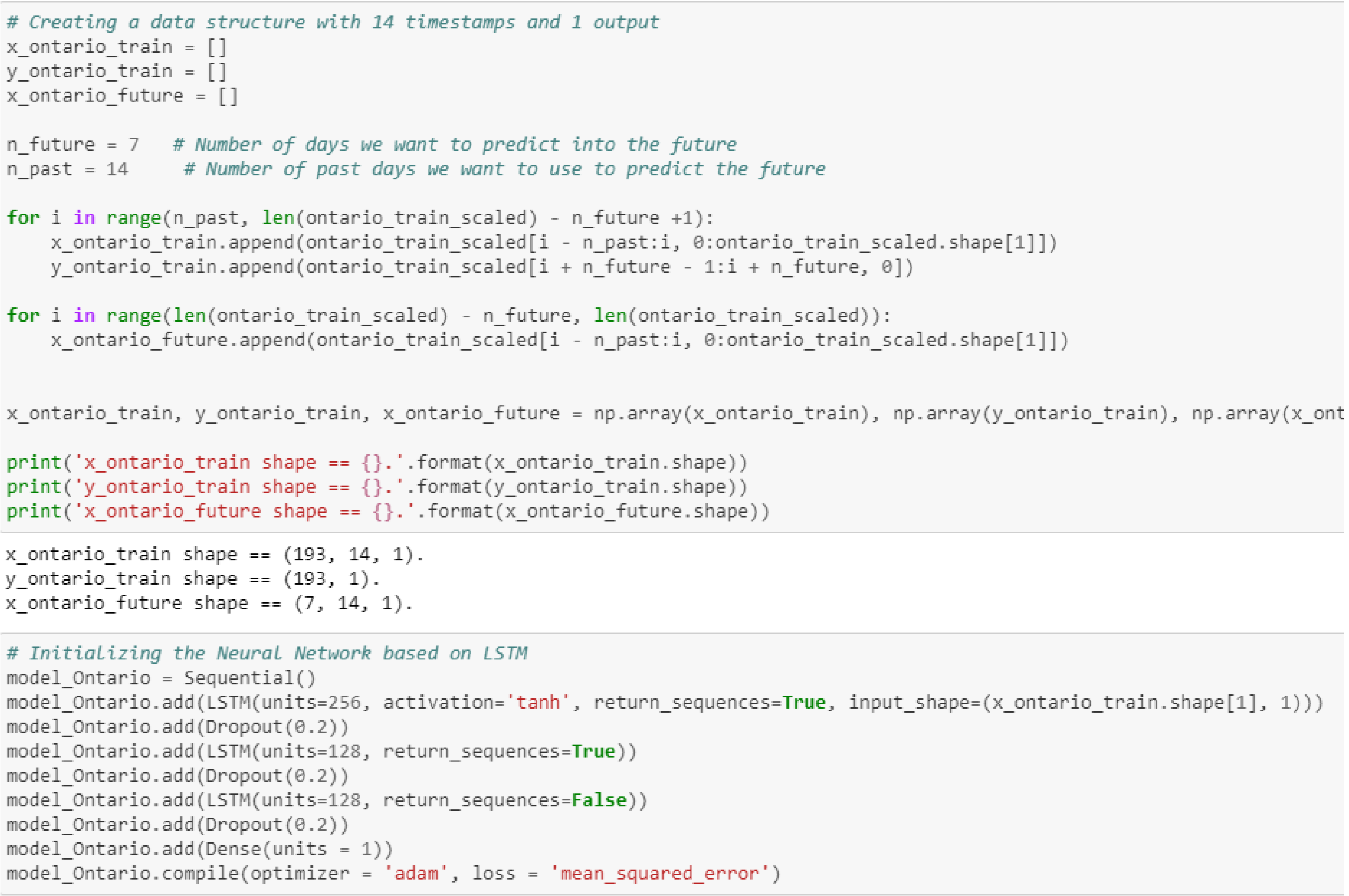
$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$

$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

LSTM Model Implementation



```
# Creating a data structure with 14 timestamps and 1 output
x_ontario_train = []
y_ontario_train = []
x_ontario_future = []

n_future = 7    # Number of days we want to predict into the future
n_past = 14      # Number of past days we want to use to predict the future

for i in range(n_past, len(ontario_train_scaled) - n_future +1):
    x_ontario_train.append(ontario_train_scaled[i - n_past:i, 0:ontario_train_scaled.shape[1]])
    y_ontario_train.append(ontario_train_scaled[i + n_future - 1:i + n_future, 0])

for i in range(len(ontario_train_scaled) - n_future, len(ontario_train_scaled)):
    x_ontario_future.append(ontario_train_scaled[i - n_past:i, 0:ontario_train_scaled.shape[1]])


x_ontario_train, y_ontario_train, x_ontario_future = np.array(x_ontario_train), np.array(y_ontario_train), np.array(x_ont

print('x_ontario_train shape == {}.'.format(x_ontario_train.shape))
print('y_ontario_train shape == {}.'.format(y_ontario_train.shape))
print('x_ontario_future shape == {}.'.format(x_ontario_future.shape))
```

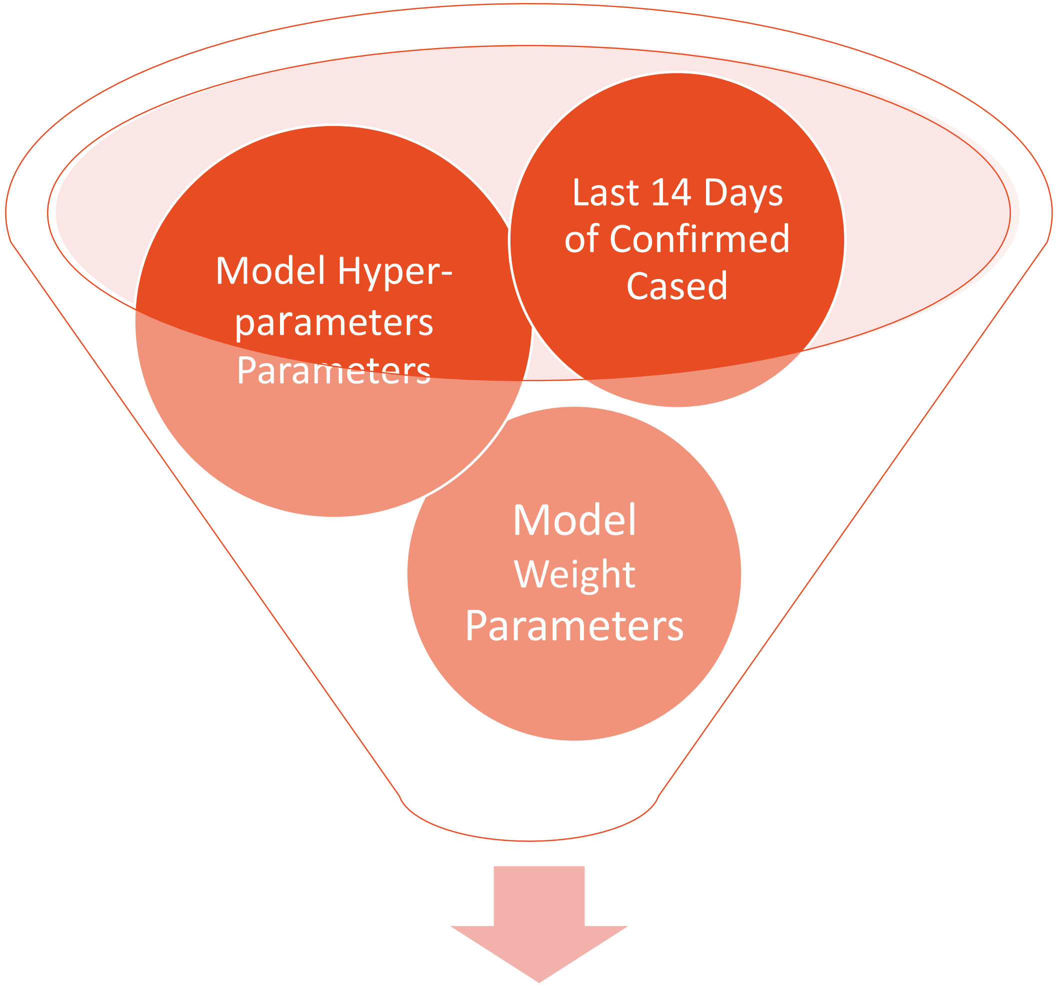
```
x_ontario_train shape == (193, 14, 1).
y_ontario_train shape == (193, 1).
x_ontario_future shape == (7, 14, 1).
```

```
# Initializing the Neural Network based on LSTM
model_Ontario = Sequential()
model_Ontario.add(LSTM(units=256, activation='tanh', return_sequences=True, input_shape=(x_ontario_train.shape[1], 1)))
model_Ontario.add(Dropout(0.2))
model_Ontario.add(LSTM(units=128, return_sequences=True))
model_Ontario.add(Dropout(0.2))
model_Ontario.add(LSTM(units=128, return_sequences=False))
model_Ontario.add(Dropout(0.2))
model_Ontario.add(Dense(units = 1))
model_Ontario.compile(optimizer = 'adam', loss = 'mean_squared_error')
```

Tools: Python, Jupyter Notebooks & IBM Cloud Park for Data
Libraries & APIs: pandas, numpy, sklearn, tensorflow.keras.LSTM, matplotlib, plotly

Method :

Training LSTM Models


Model Hyper-parameters Parameters
Last 14 Days of Confirmed Cased
Model Weight Parameters

Predict 7th day in future

Assumptions

- The last two days of reporting are ignored as they are found to be very susceptible to changes.
- Any reporting day before February 19, 2020 is ignored as there are gaps in the reporting.
- Accurate Episode Date is used to track sequence in time series.
- Any record with available Null value in any of the selected features is ignored.
- Records with “unknown” value in the “age_group” field are ignored.
- Records for last 7 days of the data are used for testing the model accuracy.

Accuracy Test

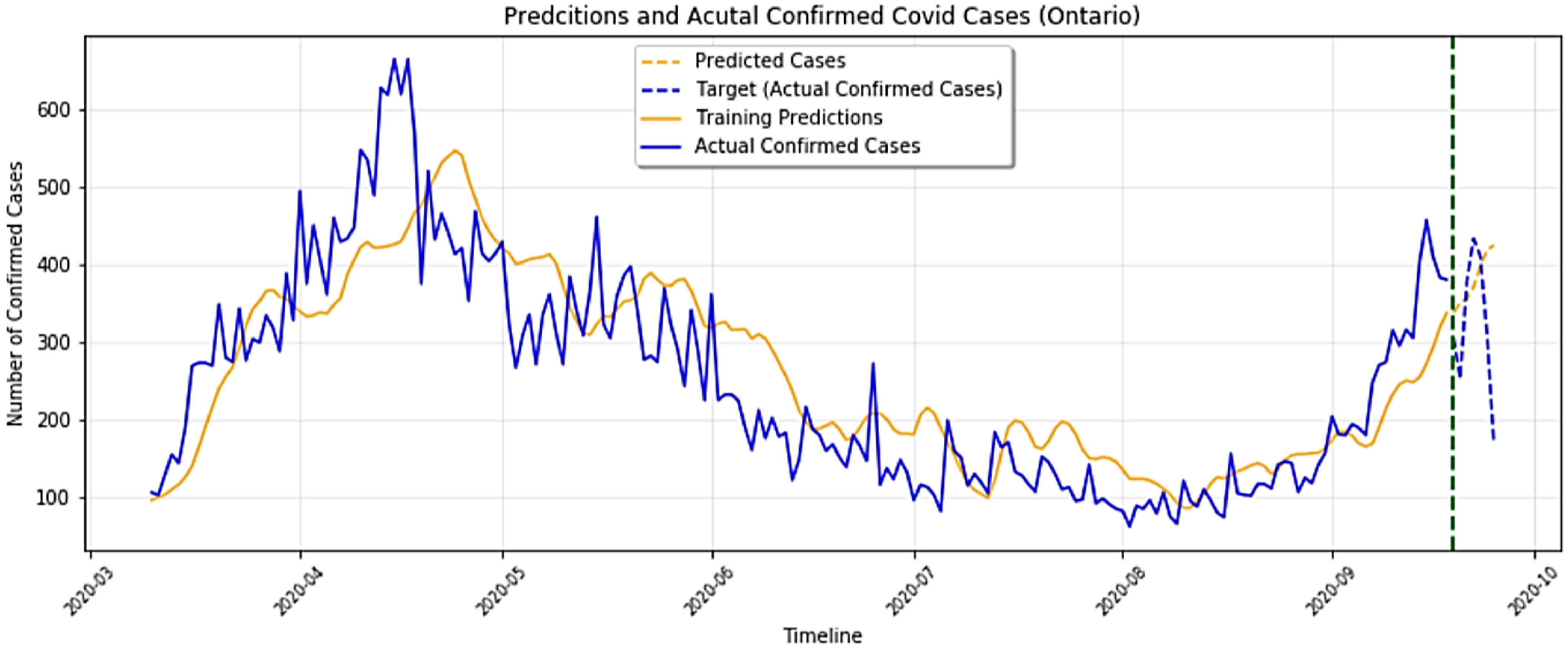
- Root-mean-square Error (RMSE) is the method used to calculate the accuracy in the prediction of models[8].

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}$$

$Y\hat{}_i$ is the i th predicted value and Y_i is the i th actual/observed value.

Results

Predictions for all of Ontario



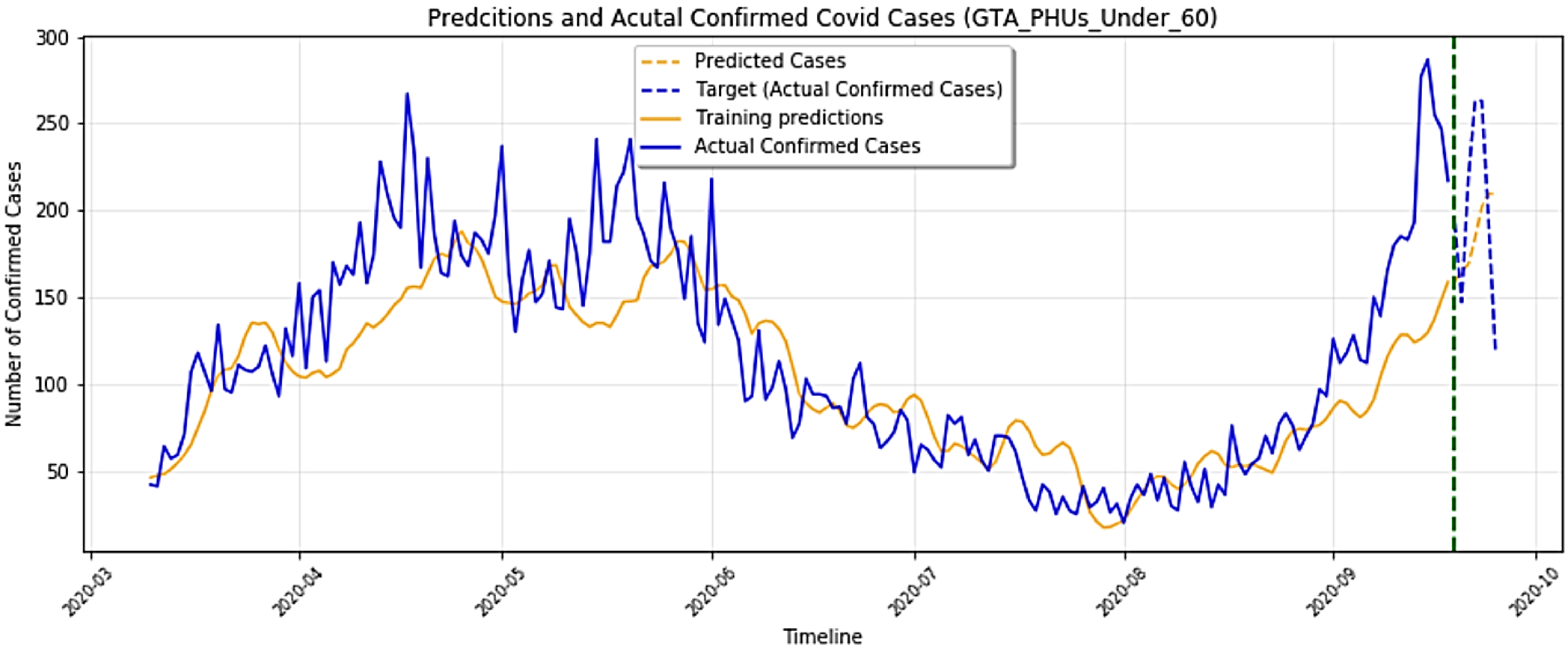
$$\text{Probability of Contracting COVID} = \frac{\text{Mean of predicted values}}{\text{Estimated Population}} * 100$$

= 379/14,864,428
≃ 0.0025 %

RMSE ≃ 115
Mean of predicted values ≃ 379
Median of predicted values ≃ 371
Standard Deviation of predicted values ≃ 32

Results

Predictions for GTA PHUs and age <60



$$\text{Probability of Contracting COVID} = \frac{\text{Mean of predicted values}}{\text{Estimated Population}} * 100$$

= 185/5,615,653

≃ 0.0032 %

RMSE ≃ 56

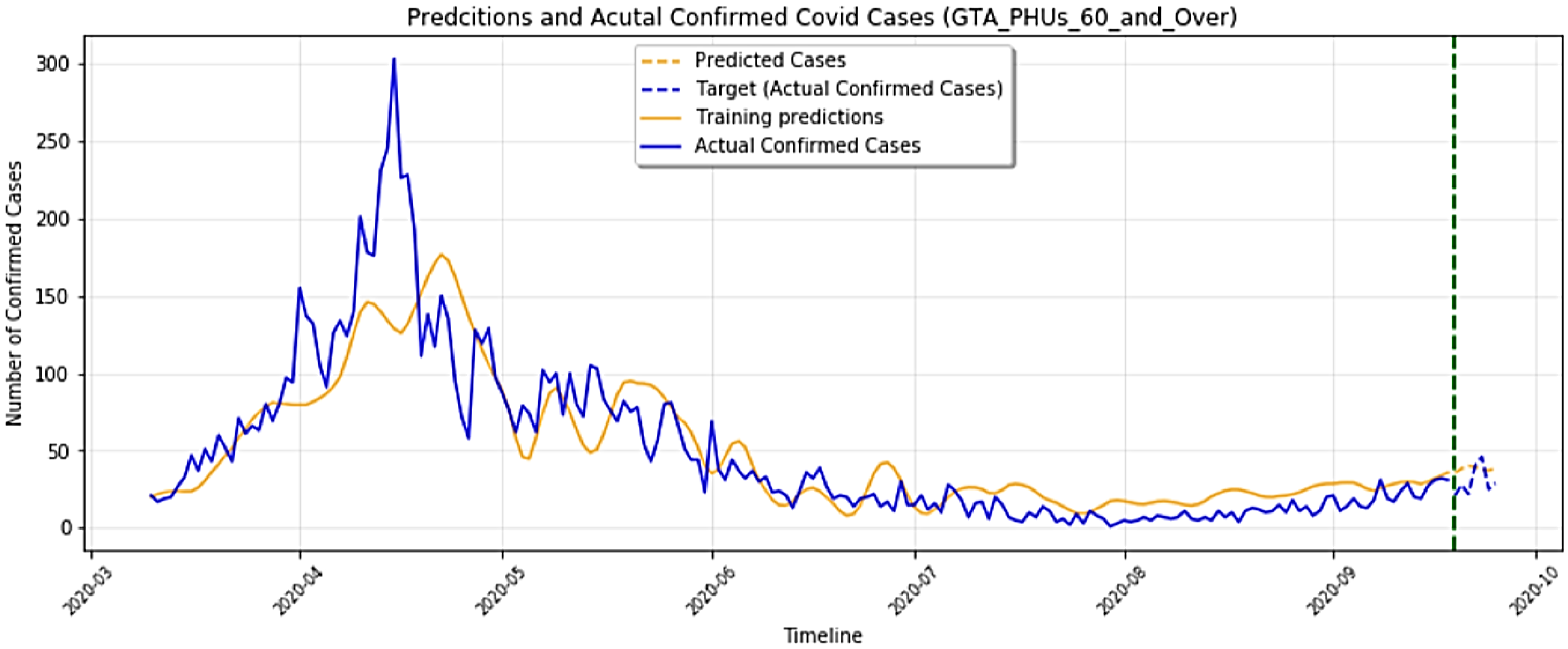
Mean of predicted values ≃ 185

Median of predicted values ≃ 184

Standard Deviation of predicted values ≃ 20

Results

Predictions for GTA PHUs and age >=60



$$\text{Probability of Contracting COVID} = \frac{\text{Mean of predicted values}}{\text{Estimated Population}} * 100$$

$$= 38/1{,}667{,}943$$

$$\simeq 0.0022\ \%$$

RMSE ≃ 12

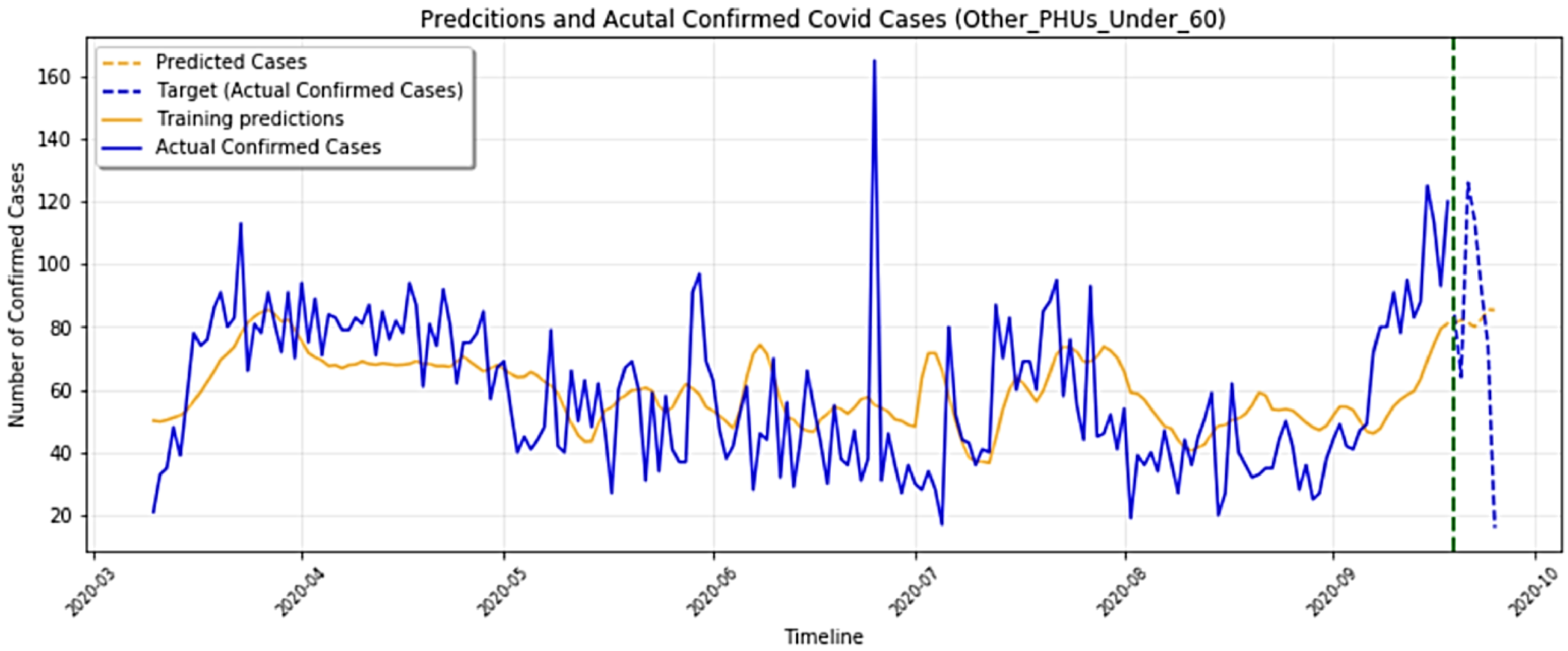
Mean of predicted values ≃ 38

Median of predicted values ≃ 38

Standard Deviation of predicted values ≃ 1

Results

Predictions for Other PHUs and age <60



$$\text{Probability of Contracting COVID} = \frac{\text{Mean of predicted values}}{\text{Estimated Population}} * 100$$

$= 83/5{,}844{,}821$

$\simeq 0.0014$ %

RMSE $\simeq$ 35

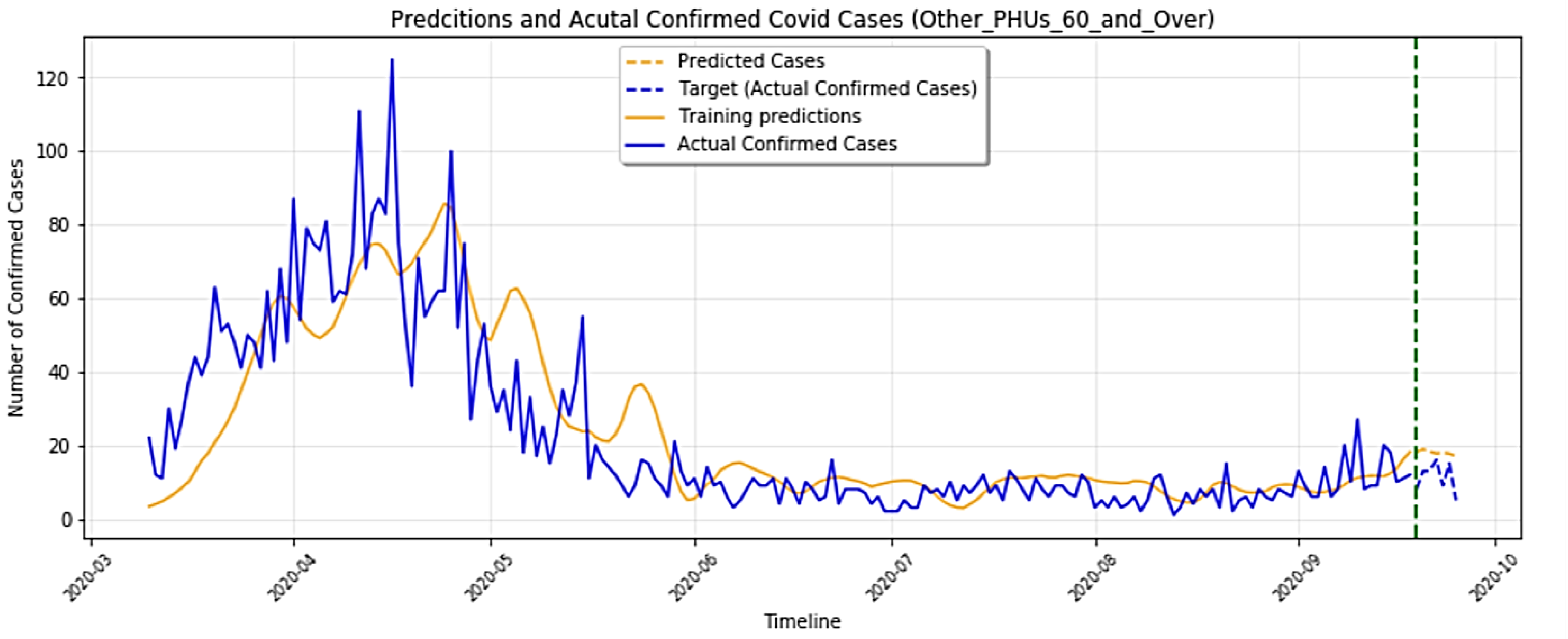
Mean of predicted values $\simeq$ 83

Median of predicted values $\simeq$ 82

Standard Deviation of predicted values $\simeq$ 2

Results

Predictions for Other PHUs and age >=60


Predcitions and Acutal Confirmed Covid Cases (Other_PHUs_60_and_Over)

Probability of Contracting COVID = $\frac{\text{Mean of predicted values}}{\text{Estimated Population}} * 100$

= 18/1,736,011

≃ 0.0010 %

RMSE ≃ 8

Mean of predicted values ≃ 18

Median of predicted values ≃ 18

Standard Deviation of predicted values ≃ 1

Results

Predictions Summary

Model Name	RMSE	RMSE/Estimated Population	Probability Result
Ontario	115	115/14,864,428 = 8*10^-6	0.0025%
GTA PHUs and age <60	56	56/5,615,653 = 10*10^-6	0.0032%
GTA PHUs and age >=60	12	12/1,667,943 = 7*10^-6	0.0022%
Other PHUs and age <60	35	35/5,844,821 = 6*10^-6	0.0014%
Other PHUs and age >=60	8	8/1,736,011 = 5*10^-6	0.0010%

Limitations

- The current models are not capable of looking into probability of contracting the virus in more detailed scenarios.
- The resulted accuracy for most models is not very high.
- The predictions are happening for 7 days into future. This can be the cause of the low model performance.

Next steps

Look into additional features (household information, outbreaks, case acquisition)

Linking to policy changes and non-pharmaceutical interventions (NPIs))

More comprehensive literature review

Testing and validating alternative models

Thank you!

References

- [1] R. M. Anderson, H. Heesterbeek, D. Klinkenberg, and T. D. Hollingsworth, “How will country-based mitigation measures influence the course of the covid-19 epidemic?” The Lancet, vol. 395, no. 10228, pp. 931–934, 2020.
- [2] A. J. Kucharski, T. W. Russell, C. Diamond, Y. Liu, J. Edmunds, S. Funk, R. M. Eggo, F. Sun, M. Jit, J. D. Munday et al., “Early dynamics of transmission and control of covid-19: a mathematical modelling study,” The lancet infectious diseases, 2020.
- [3] A. Camacho, A. Kucharski, Y. Aki-Sawyerr, M. A. White, S. Flasche, M. Baguelin, T. Pollington, J. R. Carney, R. Glover, E. Smout et al., “Temporal changes in ebola transmission in sierra leone and implications for control requirements: a real-time modelling study,” PLoS currents, vol. 7, 2015.
- [4] M. Zandavi and T. Rashidi and F. Vafaee, “Forecasting the Spread of Covid-19 Under Control Scenarios Using LSTM and Dynamic Behavioral Models,” physics.soc-ph, eprint={2005.12270}, {arXiv}, 2020.
- [5] Statistics Canada, 2016 Census of Population, Census Profile, 2016 Census. https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page.cfm?Lang=E&Geo1=PR&Code1=35&Geo2=PR&Code2=01&SearchText=35&SearchType=Begins&SearchPR=01&B1=Population&TABID=3&type=1
- [6] Public Health Ontario, Ontario COVID-19 Data Tool. https://www.publichealthontario.ca/en/data-and-analysis/infectious-disease/covid-19-data-surveillance/covid-19-data-tool
- [7] Government of Canada, People who are at high risk for severe illness from COVID-19. https://www.canada.ca/en/public-health/services/publications/diseases-conditions/people-high-risk-for-severe-illness-covid-19.html
- [8] W.Ahmed and M.Bahador, “The accuracy of the LSTM model for predicting the S&P 500 indexand the difference between prediction and backtesting,” Degree Project In Technology 2018. https://www.diva-portal.org/smash/get/diva2:1213449/FULLTEXT01.pdf

Appendix -1: Codes & Analyses

COVID-19 Exploratory Analysis:

- https://github.com/AL-DataScience/COVID-19-Predictive-Model/blob/master/COVID19_Exploratory_Analysis.ipynb

COVID-19 LSTM Predictive Model:

- https://github.com/AL-DataScience/COVID-19-Predictive-Model/blob/master/COVID%2019_LSTM_Predictive_Model.ipynb

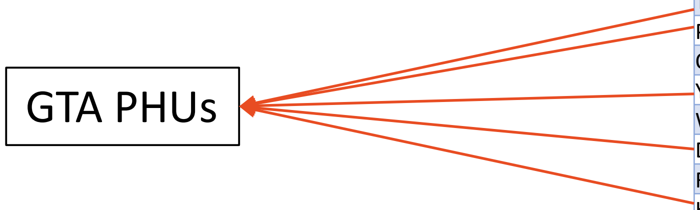
Population Analysis:

- https://github.com/AL-DataScience/COVID-19-Predictive-Model/blob/master/Population%20Analysis.xlsx

Average Confirmed Cases per Day per PHU Analysis:

- https://github.com/AL-DataScience/COVID-19-Predictive-Model/blob/master/Average_Confirmed_Cases_per_Day_per_PHU_Analysis.xlsx

Appendix -2: Average Confirmed Cases per Day


GTA PHUs

Reporting_PHU	Average Confirmed Cases per Day
Toronto Public Health	73.41
Peel Public Health	37.45
Ottawa Public Health	17.34
York Region Public Health Services	17.33
Windsor-Essex County Health Unit	11.05
Durham Region Health Department	8.88
Region of Waterloo, Public Health	7.13
Halton Region Health Department	4.73
Hamilton Public Health Services	4.72
Niagara Region Public Health Department	4.42
Simcoe Muskoka District Health Unit	3.69
Middlesex-London Health Unit	3.6
Wellington-Dufferin-Guelph Public Health	2.7
Haldimand-Norfolk Health Unit	2
Leeds, Grenville and Lanark District Health Unit	1.59
Chatham-Kent Health Unit	1.54
Lambton Public Health	1.45
Eastern Ontario Health Unit	1.11
Southwestern Public Health	1.11
Haliburton, Kawartha, Pine Ridge District Health Unit	1
Brant County Health Unit	0.83
Grey Bruce Health Unit	0.55
Kingston, Frontenac and Lennox & Addington Public Health	0.54
Huron Perth District Health Unit	0.53
Peterborough Public Health	0.5
Sudbury & District Health Unit	0.44
Thunder Bay District Health Unit	0.44
Porcupine Health Unit	0.35
Hastings and Prince Edward Counties Health Unit	0.23
Renfrew County and District Health Unit	0.22
Northwestern Health Unit	0.21
North Bay Parry Sound District Health Unit	0.16
Algoma Public Health Unit	0.13
Timiskaming Health Unit	0.07