COVID-19 Predictive Model

1 Architectural Components Overview

1.1 Data Source

Public Ontario Open Data Catalogue – Confirmed positive cases of COVID-19 in Ontario, Canada.
https://data.ontario.ca/dataset/confirmed-positive-cases-of-covid-19-in-ontario

1.1.1 Technology Choice

Python and specifically Pandas API calls are used for ETL process and feature engineering.

1.1.2 Justification

The technology is freely available.

1.2 Enterprise Data

1.2.1 Technology Choice

NA

1.2.2 Justification

No enterprise data is used and only publicly available information.

1.3 Streaming analytics

1.3.1 Technology Choice

NA

1.3.2 Justification

No streaming analytics is used.

1.4 Data Integration

1.4.1 Technology Choice

NA

1.4.2 Justification

The data and labels are available in one dataset, so no data integration is required.

1.5 Data Repository

1.5.1 Technology Choice

The downloaded data is stored on IBM cloud storage.

1.5.2 Justification

Technology is freely available.

1.6 Discovery and Exploration

1.6.1 Technology Choice

Python programing and specifically Plotly API calls are used for data exploration and creation of charts and graphs.

1.6.2 Justification

Technology is freely available and produces advanced interactive visualizations.

1.7 Actionable Insights

1.7.1 Technology Choice

Tenserflow Keras LSTM model is selected. Design: 3 LSTM Layers (256,128,128 units) followed by a Dense layer (1 unit). The accuracy of the model is measured using Root Mean Square Error (RMSE).

1.7.2 Justification

As the goal is to process timeseries information, Tenserflow Keras LSTM model is selected as the model of choice.

1.8 Applications / Data Products

1.8.1 Technology Choice

NA

1.8.2 Justification

No application is developed from this exercise.

1.9 Security, Information Governance and Systems Management

1.9.1 Technology Choice

NA

1.9.2 Justification

Not Applicable was no Application was developed.