Covid New case prediction

Assignment 2- CS4200

Professor: Hussain Zaidi

Roya Salei

rsalei@cpp.edu

December 8, 2021

# Requirement:

Design a Coronavirus Case Predictor.

• Use different markers / attributes from around the US and relate them to Coronavirus Case

Count

• Learning Model can be created using Decision Trees, Induction Learning or Neural Networks

• Model can be used to predict future case count.

# Implementation:

## Data Gathering

I have gathered U.S. Covid data from two sources:

1. Covid Timeseries Cases: https://apidocs.covidactnow.org/
2. States Mask Mandate: https://data.cdc.gov/Policy-Surveillance/U-S-State-and-Territorial-Public-Mask-Mandates-Fro/62d6-pm5i

Additionally, added the United States regions and divisions (see appendix for details) https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us\_regdiv.pdf

## Data Visualization

Time series of 51 states are included from 2020-10-06 to 2021-11-08

Preparing Data is presented in “01-Prepare-Data.ipynb” notebook

Data visualization is presented in “02-Covid-Statistics.ipynb “ notebook

## Correlation Matrix

In order to choose features that have stronger relationship with new cases of covid, I used correlation matrix.

Chart

Description automatically generated

## Models:

In this project, I used 3 different models to evaluate this data and be able to compare results with each other.

## Model 1 – Timeseries

This model is in “Model-01-timeseries.ipynb” notebook

In this model, I used TimeseriesGenerator function from tensorflow.keras.preprocessing.sequence library.

My goal was to predict new cases based on the new cases of the past 6 days. So, I set the dates as index and split the data by date range to keep the order of dates continuous.

# split data by date range

df\_train    = dff['2020-09-01': '2021-09-01']

df\_validate = dff['2021-09-02':]

In the first test, I used 64 nodes for the first layer, 64 nodes for hidden layers and 1 for the output.

tf.random.set\_seed(13)

model\_1 = tf.keras.Sequential([

tf.keras.layers.Dense(64, activation='relu', input\_dim=6),

tf.keras.layers.Dense(64, activation='relu'),

tf.keras.layers.Dense(1)

])

Model: "sequential"

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Layer (type) Output Shape Param #

=================================================================

dense (Dense) (None, 64) 448

dense\_1 (Dense) (None, 64) 4160

dense\_2 (Dense) (None, 1) 65

=================================================================

Total params: 4,673

Trainable params: 4,673

Non-trainable params: 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Chart

Description automatically generated

Figure 12 - validation loss and loss from cost functions

Chart, scatter chart

Description automatically generated

## Model 2 – Multiple Linear Regression

This model is in “Model-02-MultipleLinearRegression.ipynb” notebook

R-Squared : 0.7325632563538057

Chart

Description automatically generated

## Model 3 – Neural Network

This model is in “Model\_03\_mse\_Covid\_New\_Case\_Predictor.ipynb” notebook

Table

Description automatically generated

Chart, line chart

Description automatically generated

Figure 2 - Loss function

Graphical user interface

Description automatically generated

Chart, scatter chart

Description automatically generated

# Appendix

# Census Regions and Divisions of the United States

According to the U.S. Census Bureau United States comprise of 5 regions and 9 divisions as follow:

Table 1 Source: https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us\_regdiv.pdf

|  |  |  |
| --- | --- | --- |
| Region | Division | States |
| Region 1: Northeast | Division I: New England | Connecticut (09) Maine (23) Massachusetts (25) New Hampshire (33) Rhode Island (44) Vermont (50) |
| Division 2: Middle Atlantic | New Jersey (34) New York (36) Pennsylvania (42) |
| Region 2: Midwest | Division 3: East North Central | Indiana (18) Illinois (17) Michigan (26) Ohio ( 39) Wisconsin (55) |
| Division 4: West North Central | Iowa (19)  Nebraska (31) Kansas (20)  Minnesota (27) Missouri (29)  North Dakota (38)  South Dakota (46) |
| Region 3: South | Division 5: South Atlantic | Delaware (10)  District of Columbia (11)  Florida (12) Georgia (13) Maryland (24) North Carolina (37) South Carolina (45) Virginia (51) West Virginia (54) |
| Division 6: East South Central | Alabama (01) Kentucky (21) Mississippi (28) Tennessee (47) |
| Division 7: West South Central | Arkansas (05) Louisiana (22) Oklahoma (40) Texas (48) |
| Region 4: West | Division 8: Mountain | Arizona ( 04) Colorado (08) Idaho (16) New Mexico (35) Montana (30) Utah (49) Nevada (32) Wyoming (56) |
| Division 9: Pacific | Alaska (02) California (06) Hawaii (15) Oregon (41) Washington (53) |

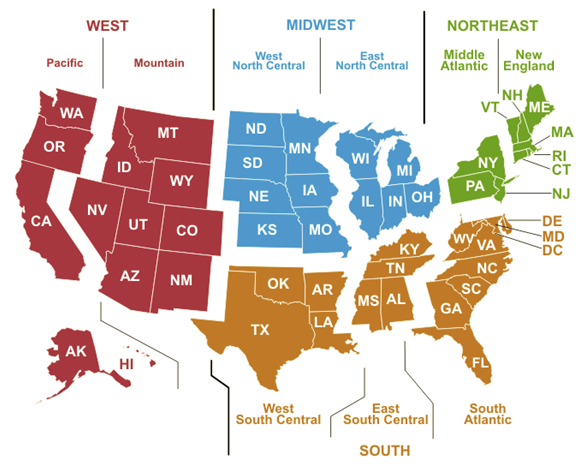


Figure 3 Census Regions and Divisions of the United States